THE ANNALS
of
MATHEMATICAL STATISTICS

(FOUNDED BY H. C. CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

Vol. XIII, No. 3 — September, 1942

THE ANNALS OF MATHEMATICAL STATISTICS

EDITED BY

S. S. WILKS, *Editor*

A. T. CRAIG J. NEYMAN

The ANNALS OF MATHEMATICAL STATISTICS is published quarterly by the Institute of Mathematical Statistics, Mt. Royal & Guilford Aves., Baltimore, Md. Subscriptions, renewals, orders for back numbers and other business communications should be sent to the ANNALS OF MATHEMATICAL STATISTICS, Mt. Royal & Guilford Aves., Baltimore, Md., or to the Secretary of the Institute of Mathematical Statistics, E. G. Olds, Carnegie Institute of Technology, Pittsburgh, Pa. Changes in mailing address which are to become effective for a given issue should be reported to the Secretary on or before the 15th of the month preceding the month of that issue. The months of issue are March, June, September and December.

Manuscripts for publication in the ANNALS OF MATHEMATICAL STATISTICS should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the ANNALS is $5.00 per year. Single copies $1.50. Back numbers are available at $5.00 per volume, or $1.50 per single issue.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC.
BALTIMORE, MD., U. S. A.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the Act of March 3, 1879

ADDITIVE PARTITION FUNCTIONS AND A CLASS OF STATISTICAL HYPOTHESES

By J. Wolfowitz

New York City

1. Introduction. The purpose of the first part of this paper is to prove several theorems about a class of functions of partitions which are additive in structure and subject to mild restrictions. These theorems may be regarded as contributions to the theory of numbers, but if one makes certain assignments of probabilities to the partitions the theorems may be expressed as statements about asymptotic distributions. It is in this latter, probabilistic language, that we shall carry out the proofs, for the following reasons. The discussion will be more concise and certain circumlocutions will be avoided. The theorems have statistical application and a number of theorems discussed recently in statistical literature are corollaries of one of our theorems.

In the second part of this paper the theory of testing statistical hypotheses where the form of the distribution functions is totally unknown and only continuity is assumed, will be discussed. The exact extension of the likelihood ratio criterion to this case will be given. Approximations to the application of this criterion in two problems will be proposed, one of which applies the results mentioned above. Lastly, in connection with the second problem, a combinatorial problem will be solved which is new and has interest per se.

2. Partitions of a single integer. Let n be a positive integer and $A = (a_1, a_2, \cdots, a_s)$ be any sequence of positive integers a_i $(i = 1, 2, \cdots, s)$, where $\sum_{i=1}^{s} a_i = n$, and s may be any integer from 1 to n. Two sequences A which have different elements or the same elements arranged in different order are to be considered distinct, so it is easy to see that there are 2^{n-1} sequences A. We shall consider the sequence A as a stochastic variable and assign to all sequences A the same probability, which is therefore 2^{-n+1}. Let r_j be the number of elements a in A which equal j $(j = 1, 2, \cdots, n)$, so that r_j is a stochastic variable. Let k be an integer $\leq n$. Then the joint distribution of the stochastic variables $r_1, r_2, \cdots, r_k$ is given as follows: The probability that $r_i = b_i$ $(i = 1, 2, \cdots, k)$ is

$$2^{-n+1}\left(\sum_{r=1}^{n} \sum \frac{r!}{(b_1)!(b_2)! \cdots (b_k)!(r_{(k+1)})! \cdots (r_n)!}\right), \tag{2.1}$$

where the inner summation is carried out over all sets of non-negative integers $r_{(k+1)}, \cdots, r_n$ such that

$$b_1 + b_2 + \cdots + b_k + r_{(k+1)} + \cdots + r_n = r, \tag{2.2}$$

$$b_1 + 2b_2 + \cdots + kb_k + (k+1)r_{(k+1)} + \cdots + nr_n = n. \tag{2.3}$$

(The b_i, of course, are non-negative integers.)

Let $r = \sum_{i=1}^{n} r_i$, and

$$r'_{(k+1)} = \sum_{i=k+1}^{n} r_i, \qquad (k < n),$$

so that r and r'_{k+1} are both stochastic variables. The probability that at the same time

$$r_i = b_i, \qquad (i = 1, \cdots, k), \tag{2.4}$$

and

$$r'_{(k+1)} = b'_{(k+1)}, \tag{2.5}$$

is given by (2.1) with the restriction

$$r_{(k+1)} + \cdots + r_n = b'_{(k+1)}, \tag{2.6}$$

added to the restrictions (2.2) and (2.3). With this added restriction the summation in (2.1) may be performed as follows: Let $t = \sum_{i=1}^{k} ib_i$. It is easy to see that the number of sequences A where every $a_i > k$, $r = r'_{(k+1)} = b'_{(k+1)}$, and $\Sigma a_i = n - t$, is given by the coefficient of x^{n-t} in the purely formal expansion in x of

$$(x^{k+1} + x^{k+2} + x^{k+3} + \cdots)^{b'_{(k+1)}} = x^{(k+1)b'_{(k+1)}} \left(\frac{1}{1-x}\right)^{b'_{k+1)}},$$

and is

$$\binom{n - t - kb'_{(k+1)} - 1}{b'_{(k+1)} - 1}.$$

Hence $P\{(2.4) \text{ and } (2.5)\}$, where this symbol will always denote the probability of the relation in braces, is seen to be

$$2^{-n+1} \frac{\left(\sum_{i=1}^{k} b_i + b'_{(k+1)}\right)!}{(b'_{(k+1)})! \prod_{i=1}^{k} (b_i)!} \binom{n - t - kb'_{(k+1)} - 1}{b'_{(k+1)} - 1}. \tag{2.7}$$

If X is a stochastic variable, let $E(X)$ and $\sigma^2(X)$ denote, respectively, the mean and variance of X (if they exist), and if Y is another stochastic variable, let $\sigma(XY)$ be the covariance between X and Y. Also let $\bar{X} = \dfrac{X - E(X)}{\sigma(X)}$. By the distribution of X we shall mean a function $\varphi(x)$ such that $P\{X < x\} \equiv \varphi(x)$. These conventions being established, we seek first to evaluate $E(r_i)$. This may be done by differentiating with respect to y the coefficient of x^n in the

purely formal expansion in x of $2^{-n+1}(x + x^2 + \cdots + x^{i-1} + yx^i + x^{i+1} + \cdots)^r$, setting $y = 1$ and summing over all values of r. We have therefore to evaluate

$$2^{-n+1} \cdot \sum_{r=2}^{n} r \binom{n-i-1}{r-2},$$

which is easily seen to give us the result

$$E(r_i) = (n - i + 3)2^{-i-1}, \qquad (i < n), \tag{2.8}$$

while it is obvious that

$$E(r_n) = 2^{-n+1}. \tag{2.9}$$

By use of similar devices the variances and covariances of the r_i may also be obtained. We omit the details of those calculations and also the presentation of the covariances, since the latter are not necessary for the proof of Theorem 2. The results are:

$$\sigma^2(r_i) = n\left(\frac{1}{2^{i+1}} + \frac{3-2i}{2^{2i+2}}\right) + \left(\frac{3-i}{2^{i+1}} + \frac{3i^2 - 12i + 5}{2^{2i+2}}\right), \qquad (i < \tfrac{1}{2}n). \tag{2.10}$$

The limitation on the value of i is necessary because the processes for summing binomial coefficients with the aid of the device described above are no longer applicable. The matter is easily settled, however, for if X is a stochastic variable which can take only the values 0 or 1, then

$$\sigma^2(X) = E(X) - [E(X)]^2.$$

The r_i for $i > \frac{1}{2}n$ are such variables, so that

$$\sigma^2(r_i) = \frac{n-i+3}{2^{i+1}} - \frac{(n-i+3)^2}{2^{2i+2}}, \qquad (n > i > \tfrac{1}{2}n), \tag{2.11}$$

$$\sigma^2(r_n) = \frac{(2^{n-1}-1)}{2^{2n-2}}. \tag{2.12}$$

Also without difficulty we have

$$\sigma^2(r_{\frac{1}{2}n}) = \frac{n+6}{2^{\frac{1}{2}(n+4)}} - \frac{(n+6)^2}{2^{n+4}} + \frac{1}{2^{n-2}}, \tag{2.13}$$

when n is even and > 2, and

$$E(r) = \tfrac{1}{2}(n+1), \tag{2.14}$$

$$\sigma^2(r) = \tfrac{1}{4}(n-1). \tag{2.15}$$

Finally,

$$E(r'_{(k+1)}) = (n - k + 1)2^{-k-1}. \tag{2.16}$$

The next results we shall need may be expressed in the following:

THEOREM 1: *As n approaches infinity, the joint distribution of the stochastic*

variables $\bar{r}_1, \cdots, \bar{r}_k, \bar{r}'_{(k+1)}$ (*k any fixed positive integer*), *approaches the multivariate normal distribution.*

This theorem is proved as follows: Make the substitutions

$$x_i = \frac{r_i - n \cdot 2^{-i-1}}{\sqrt{n}}, \qquad (i = 1, 2, \cdots, k),$$

$$x'_{(k+1)} = \frac{r'_{(k+1)} - n \cdot 2^{-k-1}}{\sqrt{n}}$$

in the expression

$$2^{-n+1} \frac{\left(\sum_{i=1}^{k} r_i + r'_{(k+1)}\right)!}{(r'_{(k+1)})! \prod_{i=1}^{k} (r_i)!} \binom{n - t - kr'_{(k+1)} - 1}{r'_{(k+1)} - 1},$$

which comes from (2.7), and regard t as equal to $\sum_{i=1}^{k} ir_i$. Replace the various factorials by their asymptotic approximations as given by Stirling's formula and simplify the resulting expression. The subsequent procedure is simple but laborious and we omit the details, which are like those of the classical proof of De Moivre's theorem as given, for example, in Frechet [1], p. 89.

We now prove the following theorem on additive partition functions:

THEOREM 2: *Let $f(x)$ be a function defined for all positive integral values of x which fulfills the following conditions*:

(a). *There exists a pair of positive integers, a and b, such that*

$$\frac{f(a)}{f(b)} \neq \frac{a}{b}, \tag{2.17}$$

(b). *the series*

$$\sum_{i=1}^{\infty} | f(i) | \, 2^{-\frac{1}{2}i}, \tag{2.18}$$

converges. Let $F(A)$, a function of the stochastic sequence A, be defined as follows:

$$F(A) = \sum_{i=1}^{s} f(a_i). \tag{2.19}$$

Then for any real y the probability of the inequality $\bar{F}(A) < y$, approaches

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{1}{2}y^2}\, dy,$$

as $n \to \infty$.

We restate this theorem without use of probabilistic terms:

Let A be any sequence of positive integers whose sum is a given integer n. Consider two sequences A to be different if they contain different elements or

the same elements arranged in a different order. Let $f(x)$ and $F(A)$ be defined as above, with the aforementioned restrictions. Then there exist, for every positive integer n, two numbers E_n and σ_n, such that 2^{-n+1} multiplied by the number of sequences A for which the inequality

$$F(A) - E_n < y\sigma_n,$$

holds, approaches

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{1}{2}y^2}\, dy,$$

as $n \to \infty$.

For convenience, the proof will be divided into a number of lemmas.

If $\varphi(y)$ is any continuous distribution function, then it is well known that $\varphi(y)$ is uniformly continuous and that consequently, for any arbitrarily small, positive ϵ, there exist two positive numbers, h and D, with the following properties:

(a). If y_1 and y_2 are any real numbers such that $|y_1 - y_2| < h$, then $|\varphi(y_1) - \varphi(y_2)| < \epsilon$,

(b). If y is such that $|y| > D$, then $\varphi(|y|) > 1 - \epsilon$, and $\varphi(-|y|) < \epsilon$.

We now first prove

LEMMA 1: *Let X and Y be two stochastic variables, both of which possess finite means and variances. Suppose that there exists a continuous distribution function $\varphi(y)$ and two small positive numbers ϵ and δ* (say $\epsilon < 1/10$, $\delta < 1/10$), *such that*

$$|P\{\bar{X} < y\} - \varphi(y)| < \epsilon, \tag{2.20}$$

for all y, and

$$\frac{\sigma(Y)}{\sigma(X)} = \delta. \tag{2.21}$$

Let h and D be chosen as above for $\varphi(y)$, with the additional proviso that $h < \frac{1}{2}$ and $D > 1$ Suppose further that

$$\delta < \min\left(\frac{h}{4D}, \frac{h\epsilon}{8}\right). \tag{2.22}$$

Then

$$|P\{(\overline{X + Y}) < y\} - \varphi(y)| < 3\epsilon, \tag{2.23}$$

for all y.

PROOF: We have

$$\sigma^2(X + Y) = \sigma^2(X) + 2\sigma(XY) + \sigma^2(Y).$$

Since, as is well known,

$$|\sigma(XY)| \le \sigma(X)\sigma(Y),$$

it follows from (2.21) that

$$\sigma(X + Y) = (1 + \delta')\sigma(X), \tag{2.24}$$

where $|\delta'| \leq \delta$ Hence

$$\sigma\left(\frac{Y - E(Y)}{\sigma(X + Y)}\right) < 2\delta. \tag{2.25}$$

From Tchebycheff's inequality and (2.21) it then follows that, if $d = h/4$,

$$P\left\{\left|\frac{Y - E(Y)}{\sigma(X + Y)}\right| > d\right\} < 4\frac{\delta^2}{d^2}, \tag{2.26}$$

and

$$\frac{4\delta^2}{d^2} < \epsilon^2 < \epsilon. \tag{2.27}$$

Now

$$\begin{aligned} P\left\{\frac{X - E(X)}{\sigma(X + Y)} < y - d\right\} &= P\left\{\frac{X - E(X)}{\sigma(X + Y)} < y - d; \left|\frac{Y - E(Y)}{\sigma(X + Y)}\right| \leq d\right\} \\ &\quad + P\left\{\frac{X - E(X)}{\sigma(X + Y)} < y - d; \left|\frac{Y - E(Y)}{\sigma(X + Y)}\right| > d\right\} \\ &< P\{(\overline{X + Y}) < y\} + \epsilon \\ &= P\left\{(\overline{X + Y}) < y; \left|\frac{Y - E(Y)}{\sigma(X + Y)}\right| \leq d\right\} \\ &\quad + P\left\{(\overline{X + Y}) < y; \left|\frac{Y - E(Y)}{\sigma(X + Y)}\right| > d\right\} + \epsilon \\ &< P\left\{\frac{X - E(X)}{\sigma(X + Y)} < y + d\right\} + 2\epsilon. \end{aligned} \tag{2.28}$$

Hence, from (2.24)

$$\begin{aligned} &P\{\bar{X} < (y - d)(1 + \delta')\} - \epsilon \\ &\qquad < P\{(\overline{X + Y}) < y\} < P\{\bar{X} < (y + d)(1 + \delta')\} + \epsilon \end{aligned} \tag{2.29}$$

and consequently, from (2.20)

$$\begin{aligned} &\varphi(y - d + y\delta' - d\delta') - 2\epsilon \\ &\qquad < P\{(\overline{X + Y}) < y\} < \varphi(y + d + y\delta' + d\delta') + 2\epsilon. \end{aligned} \tag{2.30}$$

Now if $|y| \leq 2D$, then from (2.22)

$$d + |y\delta'| + d|\delta'| < \frac{h}{4} + \frac{h}{2} + \frac{h}{4} = h,$$

and if $|y| > 2D$, then also from (2.22)

$$|y| - d - |y\delta'| - d|\delta'| > |y|(1 - \delta) - \frac{h}{2} > \frac{3}{4}|y| > \frac{3}{2}D.$$

Recalling the definitions of h and D, it follows from (2.30) that, for all y,

$$\varphi(y) - 3\epsilon < P\{(\overline{X + Y}) < y\} < \varphi(y) + 3\epsilon. \tag{2.31}$$

This proves Lemma 1.

LEMMA 2: *For any fixed pair a, b, of positive integers such that $a < b$,*

$$\lim_{n\to\infty} \frac{[E(r)]^{b-a} \cdot [E(r_b)]^a}{[E(r_a)]^b} = 1 \tag{2.32}$$

PROOF: From (2.8), for fixed i

$$\frac{1}{n} E(r_i) \to 2^{-i-1},$$

and from (2.14) $\frac{1}{n} E(r) \to \frac{1}{2}$ as $n \to \infty$. The required result follows easily.

For any n we now define

$$B(k;n) = \sum_{i=1}^{k} r_i[f(i)],$$

and

$$C(k;n) = \sum_{i=k+1}^{n} r_i[f(i)].$$

Then

$$F(A) = B(k;n) + C(k;n).$$

LEMMA 3: *For any real y and any fixed positive integral k the probability that the stochastic variable $B(k;n)$ shall fulfill the inequality $\bar{B}(k;n) < y$ approaches $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{1}{2}y^2}\, dy$, as $n \to \infty$.*

PROOF: By Theorem 1, the stochastic variables $\bar{r}_1, \bar{r}_2, \cdots, \bar{r}_k, \bar{r}'_{(k+1)}$, are asymptotically jointly normally distributed. As an immediate consequence so are the variables $\bar{r}_1, \bar{r}_2, \cdots, \bar{r}_k$, and hence $B(k;n)$, which is a linear function with constant coefficients $f(1), f(2), \cdots f(k)$, of $r_1, r_2, \cdots, r_k$, is asymptotically normally distributed.

LEMMA 4. *There exists a constant $c > 0$, such that, for all n sufficiently large,*

$$\sigma^2(F(A)) > cn. \tag{2.33}$$

PROOF: For any sufficiently large, arbitrary, but fixed n, we will construct two sets, S_1 and S_2, of sequences A, with the following properties: S_1 and S_2 have the same probability p, with p always greater than β, a fixed positive

constant which does not depend on n. Since the probabilities of S_1 and S_2 are equal, each possesses the same number of sequences A. Between the member sequences of the sets S_1 and S_2 we will establish a one-to-one correspondence such that, if A_1 is a member of S_1 and A_2 is its corresponding sequence in S_2, then

$$(2.34) \qquad | F(A_1) - F(A_2) | > 2d\sqrt{n},$$

where d is a fixed positive constant which does not depend on n.

It is easy to see that such a construction would prove the lemma. The probability of any sequence A is 2^{-n+1}. Hence the contribution of a corresponding pair A_1 and A_2 to the variance of $F(A)$ is by (2.34) not less than $2^{-n+2}d^2n$ and the contribution of the sets S_1 and S_2 is not less than $2\beta d^2 n$.

It remains then to carry out the construction of S_1 and S_2. For the sake of simplicity in notation, we shall carry out the construction with the assumption that the integers a and b of (2.17) are 1 and 2. It will be readily apparent, however, that the proof is perfectly general and with trivial changes holds for any pair a, b. This lemma is the only place where the hypothesis (2.17) is used. The latter condition is necessary because, if for every pair of positive integers i and j,

$$\frac{f(i)}{f(j)} = \frac{i}{j},$$

then $F(A)$ is a constant multiple of n, for $n = \sum_i ir_i$ and then

$$F(A) = \sum_i f(a_i) = \sum_i r_i f(i) = f(1) \sum_i ir_i = nf(1).$$

Each sequence A uniquely determines the "coördinate" complex

$$\{r_1, r_2, \cdots, r_n\}$$

which we prefer to write as the pair $L = (l, l')$:

$$l = \{r_1, r_2\},$$

$$l' = \{r_3, r_4, \cdots, r_n\}.$$

To each pair (l, l') there correspond in general many sequences A whose exact number may be explicitly given in terms of factorials. The totality of all A whose L have the same second member l' will be called the group determined by l', or just the group l'. The subset of a group l' all of whose A have the same r_1 will be called the family (l', r_1). All the A in the same family have the same L. For l' and r_1 determine r_2 through the equation $\sum_i ir_i = n$.

According to Theorem 1 for $k = 2$, r_1, r_2, r_3' are asymptotically jointly normally distributed. Let

$$\sigma_1 = \lim_{n\to\infty} \frac{\sigma(r_1)}{\sqrt{n}}$$

The limiting variances of r_2 and r_3' are constant multiples of $n\sigma_1^2$. Therefore the set H of all A whose L satisfy the constraints

$$\begin{aligned} &\frac{n}{4} < r_1 < \frac{n}{4} + \sqrt{n}\,\sigma_1 \\ (2.35) \qquad &\frac{n}{8} < r_2 < \frac{n}{8} + \sqrt{n}\,\sigma_1 \\ &\frac{n}{8} < r_3' < \frac{n}{8} + \sqrt{n}\,\sigma_1 \end{aligned}$$

has, by virtue of the fact that the limiting correlation coefficients of the variables r_1, r_2, r_3' are all less than 1 in absolute value, a positive probability, which exceeds a fixed positive constant γ for sufficiently large n. If any member sequence A of a family is in H, the entire family is obviously in H. Any sequence A belongs to one and only one family. Hence the set H may be decomposed in a disjunct way into entire families. Let $\left(l', \frac{n}{4} + h_1\right)$ be *any* family in H, where of course $0 < h_1 < \sqrt{n}\,\sigma_1$. Consider the (second) family $\left(l', \frac{n}{4} + 2\sqrt{n}\,\sigma_1 + h_1\right)$. This family is not in H. We now wish to show that the probability of the second family exceeds c' times the probability of the first family, where c' is a fixed positive constant which does not depend on either n or the particular families in question.

For the first family, let

$$r_1 = \frac{n}{4} + h_1, \qquad r_3' = \frac{n}{8} + h_3,$$

$$r_2 = \frac{n}{8} + h_2, \qquad r = \frac{n}{2} + h_1 + h_2 + h_3.$$

Hence

$$(2.36) \qquad 0 < h_i < \sqrt{n}\sigma_1 \qquad (i = 1, 2, 3).$$

For the second family we therefore have, since both families are in the same group,

$$\begin{aligned} r_1 &= \frac{n}{4} + 2\sqrt{n}\,\sigma_1 + h_1, \\ r_2 &= \frac{n}{8} - \sqrt{n}\,\sigma_1 + h_2, \\ r_3' &= \frac{n}{8} + h_3, \\ r &= \frac{n}{2} + \sqrt{n}\,\sigma_1 + h_1 + h_2 + h_3. \end{aligned}$$

The ratio of the probability of the second family to that of the first family equals the ratio of the number of sequences A in the second family to the number of sequences A in the first family. By elementary combinatorics, since both families are in the same group, the latter ratio is

$$\frac{\left(\frac{n}{2} + \sqrt{n}\,\sigma_1 + h_1 + h_2 + h_3\right)!}{\left(\frac{n}{4} + 2\sqrt{n}\,\sigma_1 + h_1\right)!\left(\frac{n}{8} - \sqrt{n}\,\sigma_1 + h_2\right)!} \frac{\left(\frac{n}{4} + h_1\right)!\left(\frac{n}{8} + h_2\right)!}{\left(\frac{n}{2} + h_1 + h_2 + h_3\right)!} \tag{2.37}$$

and hence exceeds

$$\begin{aligned}&\left(\frac{n}{2} + h_1 + h_2 + h_3\right)^{\sqrt{n}\,\sigma_1} \\ &\qquad\times\left(\frac{n}{4} + 2\sqrt{n}\,\sigma_1 + h_1\right)^{-2\sqrt{n}\,\sigma_1}\left(\frac{n}{8} - \sqrt{n}\,\sigma_1 + h_2\right)^{\sqrt{n}\,\sigma_1}.\end{aligned} \tag{2.38}$$

At this point, if we had been using the numbers a and b of (2.17), we would make use of Lemma 2. In the present case the result of that lemma is trivial. It is easy to see, therefore, that (2.38) equals

$$\begin{aligned}&\left(1 + \frac{2h_1 + 2h_2 + 2h_3}{n}\right)^{\sqrt{n}\,\sigma_1} \\ &\qquad\times\left(1 + \frac{8\sqrt{n}\,\sigma_1 + 4h_1}{n}\right)^{-2\sqrt{n}\,\sigma_1}\left(1 - \frac{8\sqrt{n}\,\sigma_1 - 8h_2}{n}\right)^{\sqrt{n}\,\sigma_1},\end{aligned} \tag{2.39}$$

which, in view of (2.36), exceeds

$$\left(1 + \frac{12\sigma_1}{\sqrt{n}}\right)^{-2\sqrt{n}\,\sigma_1} \cdot \left(1 - \frac{8\sigma_1}{\sqrt{n}}\right)^{\sqrt{n}\,\sigma_1} \tag{2.40}$$

which, in turn, for sufficiently large n, exceeds

$$\tfrac{1}{2}\cdot e^{-24\sigma_1^2 - 8\sigma_1^2} = \tfrac{1}{2}e^{-32\sigma_1^2} = c'. \tag{2.41}$$

We are now ready to construct S_1 and S_2. Let

$$f_1 = (l', r_1)$$

be any family in H and consider the family

$$f_2 = (l', r_1 + 2\sqrt{n}\,\sigma_1).$$

Select in any manner whatsoever $c'\nu$ of the sequences A in f_1, where ν is the total number of sequences in f_1. Call this set of sequences f^*. Select in any manner whatsoever $c'\nu$ sequences from f_2 and call this set f^{**}. That there exist at least $c'\nu$ sequences in f_2 is assured by equation (2.41). In any manner whatsoever establish a one-to-one correspondence between the sequences of f^* and f^{**}. Suppose A_1 and A_2 are corresponding sequences. Since f^* and f^{**} belong to the same group, and since $f(2) \neq 2f(1)$, we have

$$\begin{aligned} (2.42) \qquad | F(A_1) - F(A_2) | &= | f(2)\sqrt{n}\,\sigma_1 - 2f(1)\sqrt{n}\,\sigma_1 | \\ &= | f(2) - 2f(1) | \sqrt{n}\,\sigma_1 , \end{aligned}$$

so that (2.34) holds with

$$(2.43) \qquad d = \tfrac{1}{4}\, | f(2) - 2f(1) |\, \sigma_1 .$$

Now proceed in this manner for all the families f_1 in H. The union of all the sets f^* is the set S_1 and the union of all the sets f^{**} is the set S_2. It is clear that, since the probability of H exceeds γ, the probability p of S_1 exceeds $\beta = c'\gamma$. This proves Lemma 4.

LEMMA 5. *For any arbitrarily small positive number ξ there exists a positive integer $\mu(\xi)$, such that for any $k > \mu(\xi)$ and all n greater than a fixed lower bound,*

$$(2.44) \qquad \sigma^2[C(k;n)] < \xi n.$$

PROOF: Since

$$C(k;n) = \sum_{i=k+1}^{n} r_i f(i),$$

and, as is well known,

$$| \sigma(XY) | \leq \sigma(X)\sigma(Y)$$

we have

$$(2.45) \qquad \sigma^2[C(k;n)] \leq \left[\sum_{i=k+1}^{n} | f(i) |\, \sigma(r_i) \right]^2$$

From (2.10) it follows readily that

$$(2.46) \qquad \sigma^2(r_i) < \frac{n}{2^i} + \frac{5}{2^{i+1}} + \left(\frac{-i}{2^{i+1}} + \frac{3i^2}{2^{2i+2}} \right),$$

and the quantity in parentheses in the right member of (2.46) is easily seen to be negative, so that, for $i < \frac{1}{2}n$ and $n \geq 3$,

$$(2.47) \qquad \sigma(r_i) < \sqrt{2n}\, 2^{-\frac{1}{2}i}$$

From (2.11) and the definition of r_i, it follows easily that (2.47) holds also when $i > \frac{1}{2}n$ and $n \geq 3$.

Hence, in view of (2.12), (2.13), and the convergence of the series in (2.18), the desired result follows from (2.45).

LEMMA 6. *Let the ξ of Lemma 5 be $< \frac{1}{4}c$, where c is as in Lemma 4. Then for $k > \mu(\xi)$ and n larger than a fixed lower bound*

$$(2.48) \qquad \sigma^2(B(k;n)) > \tfrac{1}{4}cn.$$

PROOF: Since

$$F(A) = B(k;n) + C(k;n),$$

we have

$$\begin{aligned}\sigma^2(F(A)) &= \sigma^2(B(k;n)) + \sigma^2(C(k;n)) + 2\sigma(BC)\\ &\leq \sigma^2(B) + \sigma^2(C) + 2\sigma(B)\sigma(C) = (\sigma(B) + \sigma(C))^2.\end{aligned}$$

Hence from (2.33) and (2.44) $\sqrt{cn} < \sigma(B) + \frac{1}{2}\sqrt{cn}$ and the required result follows.

Proof of the Theorem: Let ϵ be an arbitrarily small positive number. For all n sufficiently large we have, by Lemma 3,

$$\left| P\{\bar{B}(k;n) < y\} - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{1}{2}y^2}\,dy \right| < \epsilon,$$

for all y. For a small ξ to be chosen later and large enough k and n we have, by Lemmas 5 and 6,

$$\frac{\sigma(C(k;n))}{\sigma(B(k;n))} = \delta < \frac{4\xi}{c}. \tag{2.49}$$

Now let the $\varphi(y)$ of Lemma 1 be defined as

$$\varphi(y) \equiv \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{1}{2}y^2}\,dy,$$

and choose h and D as in Lemma 1 for our present ϵ. Since c is fixed and ξ still at our disposal, choose ξ sufficiently small so that the δ of (2.49) satisfies (2.22). Since the hypothesis of Lemma 1 is satisfied, we have, from (2.23) and Lemma 3, for all n sufficiently large,

$$| P\{\bar{F}(A) < y\} - \varphi(y) | < 3\epsilon$$

for all y. This is the required result.

3. Partitions of two integers. Let n_1 and n_2 be positive integers, $n_1 + n_2 = n$, $\frac{n_1}{n} = e_1$, $\frac{n_2}{n} = e_2$, and $e = \max(e_1, e_2)$. Let $V = (v_1, v_2, \cdots, v_s)$ be any sequence of positive integers v_i $(i = 1, 2, \cdots, s)$ where $a_1 + a_3 + a_5 + \cdots$ equals either one of n_1 and n_2, while $a_2 + a_4 + a_6 + \cdots$ equals the other. Such sequences are of statistical importance (cf. Wald and Wolfowitz [2]). As before, sequences V with different elements or with the same elements in different order will be considered different and to each sequence V will be assigned the same probability, which is therefore easily seen to be $\frac{n_1!\,n_2!}{n!}$.

Let r_{1i} be the number of elements equal to i in that one of the two sequences $(a_1, a_3, a_5, \cdots)$ and $(a_2, a_4, a_6, \cdots)$ the sum of whose elements is n_1 and let r_{2i} be the corresponding number for the other sequence. Let

$$s_i = r_{1i} + r_{2i},$$
$$r_1 = \sum_i r_{1i}, \qquad r_2 = \sum_i r_{2i},$$
$$s = r_1 + r_2, \qquad r'_{1(k+1)} = \sum_{i=k+1}^{n_1} r_{1i}$$
$$r'_{2(k+1)} = \sum_{i=k+1}^{n_2} r_{2i}.$$

The necessary computations such as are given in the beginning of the previous section have been performed by Mood [3] and we summarize them as follows:

THEOREM 3 (Mood): *As n approaches infinity while e_1 and e_2 remain constant, the joint distribution of the stochastic variables*

$$\bar{r}_{11}, \bar{r}_{12}, \cdots, \bar{r}_{1k}, \bar{r}'_{1(k+1)}, \bar{r}_{21}, \bar{r}_{22}, \cdots \bar{r}_{2k}$$

(where k is any fixed positive integer), approaches the multivariate normal distribution.

Mood (loc. cit.) gives the following parameters, with the convention that

$$x^{(i)} = x(x-1)(x-2)\cdots(x-i+1): \tag{3.1}$$

$$E(r_{1i}) = \frac{(n_2+1)^{(2)} n_1^{(i)}}{n^{(i+1)}}, \tag{3.2}$$

$$\lim_{n\to\infty} \frac{E(r_{1i})}{n} = e_1^i e_2^2, \tag{3.3}$$

$$\lim_{n\to\infty} \frac{E(r'_{1(k+1)})}{n} = e_1^{k+1} e_2, \tag{3.4}$$

$$\sigma^2(r_{1i}) = \frac{n_2^{(2)}(n_2+1)^{(2)} n_1^{(2i)}}{n^{(2i+2)}} + \frac{(n_2+1)^{(2)} n_1^{(i)}}{n^{(i+1)}}\left(1 - \frac{(n_2+1)^{(2)} n_1^{(i)}}{n^{(i+1)}}\right), \tag{3.5}$$

$$\lim_{n\to\infty} \frac{\sigma^2(r_{1i})}{n} = e_1^{2i-1} e_2^3 [(i+1)^2 e_1 e_2 - i^2 e_2 - 2e_1] + e_1^i e_2^2. \tag{3.6}$$

The corresponding parameters for r_{2i} may be obtained from the above by interchange of n_1 and n_2. Also

$$\lim_{n\to\infty} \frac{E(r_1)}{n} = \lim_{n\to\infty} \frac{E(r_2)}{n} = e_1 e_2. \tag{3.7}$$

For additive partition functions we have the following theorem:

THEOREM 4. *Let $f(x)$ be a function defined for all positive integral values of x which fulfills the following conditions:*

a) *There exists a pair of positive integers, a and b, such that*

$$\frac{f(a)}{f(b)} \neq \frac{a}{b}; \tag{3.8}$$

b) *the series*

$$\sum_{i=1}^{\infty} | f(i) | e^{i/2} \tag{3.9}$$

converges. Let $F(V)$, a function of the stochastic sequence V, be defined as follows:

$$F(V) = \sum_{i=1}^{s} f(v_i) \tag{3.10}$$

Then for any real y the probability of the inequality $\bar{F}(V) < y$ approaches

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{1}{2}y^2}\, dy,$$

as $n \to \infty$, while e_1 and e_2 remain constant.

The basic idea of the proof of this theorem is the same as that of the proof of Theorem 2. We omit all the steps which can be written without difficulty by analogy to those in Theorem 2 and present only those where some major change is necessary. The numbering of the lemmas will correspond to that of Theorem 2.

Lemma 2. *For any fixed pair, a and b, of positive integers such that $a < b$,*

$$[E(r_1)]^{b-a} \cdot [E(r_2)]^{b-a} \cdot [E(r_{1b})]^{a} \cdot [E(r_{2b})]^{a} \cdot [E(r_{1a})]^{-b} \cdot [E(r_{2a})]^{-b} \to 1, \tag{3.11}$$

as $n \to \infty$.

The proof is the same as before.

The following are the definitions corresponding to those of Theorem 2:

$$B(k;n) = \sum_{i=1}^{k} s_i f(i),$$

$$C(k;n) = \sum_{i=k+1}^{n} s_i f(i).$$

Then as before

$$F(V) = B(k;n) + C(k;n).$$

Lemma 4. Statement is the same as that for Theorem 2. The following important changes must be made in the proof:

Each sequence V determines the coordinate complex

$$\begin{Bmatrix} r_{11}, r_{12}, \cdots, r_{1n} \\ r_{21}, r_{22}, \cdots, r_{2n} \end{Bmatrix}$$

also

$$l = \begin{Bmatrix} r_{11}, r_{12} \\ r_{21}, r_{22} \end{Bmatrix},$$

and

$$l' = \begin{Bmatrix} r_{13}, & \cdots, & r_{1n} \\ r_{23}, & \cdots, & r_{2n} \end{Bmatrix}.$$

The set H is the set of all V whose L satisfy the constraints

$$\begin{aligned} ne_1e_2^2 &< r_{11} < ne_1e_2^2 + \sqrt{n}\,\sigma_{11}, \\ ne_1^2e_2^2 &< r_{12} < ne_1^2e_2^2 + \sqrt{n}\,\sigma_{11}, \\ ne_1^2e_2 &< r_{21} < ne_1^2e_2 + \sqrt{n}\,\sigma_{11}, \\ ne_1^2e_2^2 &< r_{22} < ne_1^2e_2^2 + \sqrt{n}\,\sigma_{11}, \\ ne_1^3e_2 &< r'_{13} < ne_1^3e_2 + \sqrt{n}\,\sigma_{11}, \end{aligned}$$

where

$$\sigma_{11} = \lim_{n\to\infty} \frac{\sigma(r_{11})}{\sqrt{n}}.$$

The representative family for H is characterized by

$$(l', ne_1e_2^2 + h_{11}),$$

and this family is compared with the family

$$(l', ne_1e_2^2 + 2\sqrt{n}\,\sigma_{11} + h_{11}).$$

For the members of the family in H

$$\begin{aligned} r_{11} &= ne_1e_2^2 + h_{11} = nm_{11} + h_{11}, \\ r_{12} &= ne_1^2e_2^2 + h_{12} = nm_{12} + h_{12}, \\ r_{21} &= ne_1^2e_2 + h_{21} = nm_{21} + h_{21}, \\ r_{22} &= ne_1^2e_2^2 + h_{22} = nm_{22} + h_{22}, \\ r'_{13} &= ne_1^3e_2 + h_{13} = nm'_{13} + h_{13}, \\ r_1 &= ne_1e_2 + h = nm + h, \end{aligned}$$

$$| r_2 - r_1 | \leq 1,$$

where

$$h_{ij} < \sqrt{n}\,\sigma_{11}, \tag{3.12}$$

$$h = h_{11} + h_{12} + h_{13}. \tag{3.13}$$

And for the members of the second family

$$\begin{aligned} r_{11} &= nm_{11} + 2\sqrt{n}\,\sigma_{11} + h_{11}, \\ r_{12} &= nm_{12} - \sqrt{n}\,\sigma_{11} + h_{12}, \end{aligned}$$

$$\begin{aligned} r_{21} &= nm_{21} + 2\sqrt{n}\,\sigma_{11} + h_{21} + \theta_{21}, \\ r_{22} &= nm_{22} - \sqrt{n}\,\sigma_{11} + h_{22} + \theta_{22}, \\ r'_{13} &= nm'_{13} + h_{13}, \\ r_1 &= nm + \sqrt{n}\,\sigma_{11} + h, \\ &| r_2 - r_1 | \leq 1, \end{aligned}$$

with

$$| \theta_{21} | \leq 1, \qquad | \theta_{22} | \leq 1.$$

To the expression (2.37) corresponds the expression (3.14), with $| \theta | \leq 1$:

$$\begin{aligned} &\frac{(nm_{11} + h_{11})!\,(nm_{12} + h_{12})!}{(nm + h)!} \times \frac{(nm_{21} + h_{21})!\,(nm_{22} + h_{22})!}{(nm + h)!} \\ &\times \frac{(nm + h + \sqrt{n}\,\sigma_{11})!}{(nm_{11} + 2\sqrt{n}\,\sigma_{11} + h_{11})!\,(nm_{12} - \sqrt{n}\,\sigma_{11} + h_{12})!} \\ &\times \frac{(nm + h + \sqrt{n}\,\sigma_{11} + \theta)!}{(nm_{21} + 2\sqrt{n}\,\sigma_{11} + h_{21} + \theta_{21})!\,(nm_{22} - \sqrt{n}\,\sigma_{11} + h_{22} + \theta_{22})!}, \end{aligned} \tag{3.14}$$

which exceeds

$$\begin{aligned} (nm + h)^{2\sqrt{n}\,\sigma_{11}} &\times (nm_{11} + 2\sqrt{n}\,\sigma_{11} + h_{11})^{-2\sqrt{n}\,\sigma_{11}} \\ &\times (nm_{12} - \sqrt{n}\,\sigma_{11} + h_{12})^{\sqrt{n}\,\sigma_{11}} \\ &\times (nm_{21} + 2\sqrt{n}\,\sigma_{11} + h_{21})^{-2\sqrt{n}\,\sigma_{11}} \\ &\times (nm_{22} - \sqrt{n}\,\sigma_{11} + h_{22})^{\sqrt{n}\,\sigma_{11}}. \end{aligned} \tag{3.15}$$

Employing Lemma 2, we find that (3.15) equals

$$\begin{aligned} \left(1 + \frac{h}{nm}\right)^{2\sqrt{n}\,\sigma_{11}} &\times \left(1 + \frac{2\sqrt{n}\,\sigma_{11} + h_{11}}{nm_{11}}\right)^{-2\sqrt{n}\,\sigma_{11}} \\ &\times \left(1 + \frac{-\sqrt{n}\,\sigma_{11} + h_{12}}{nm_{12}}\right)^{\sqrt{n}\,\sigma_{11}} \\ &\times \left(1 + \frac{2\sqrt{n}\,\sigma_{11} + h_{21}}{nm_{21}}\right)^{-2\sqrt{n}\,\sigma_{11}} \\ &\times \left(1 + \frac{-\sqrt{n}\,\sigma_{11} + h_{22}}{nm_{22}}\right)^{\sqrt{n}\,\sigma_{11}}. \end{aligned} \tag{3.16}$$

In view (3.12) and (3.13), (3.16) exceeds

$$\begin{aligned} &\left(1 + \frac{3\sqrt{n}\,\sigma_{11}}{nm_{11}}\right)^{-2\sqrt{n}\,\sigma_{11}} \times \left(1 - \frac{\sqrt{n}\,\sigma_{11}}{nm_{12}}\right)^{\sqrt{n}\,\sigma_{11}} \\ &\times \left(1 + \frac{3\sqrt{n}\,\sigma_{11}}{nm_{21}}\right)^{-2\sqrt{n}\,\sigma_{11}} \times \left(1 - \frac{\sqrt{n}\,\sigma_{11}}{nm_{22}}\right)^{\sqrt{n}\,\sigma_{11}}, \end{aligned} \tag{3.17}$$

which, for sufficiently large n, in turn exceeds

$$\tfrac{1}{2}\cdot e^{-\sigma_{11}^2}\left(\frac{6}{m_{11}} + \frac{1}{m_{12}} + \frac{6}{m_{21}} + \frac{1}{m_{22}}\right) = c'. \tag{3.18}$$

LEMMA 5. Statement is the same as for Theorem 2. The proof then proceeds as follows:

We have

$$\sigma^2(C(k;n)) \leq \left(\sum_{i=1}^{2}\sum_{j=k+1}^{n} |f(j)|\,\sigma(r_{ij})\right)^2. \tag{3.19}$$

From an examination of (3.5) and (3.6) we may see without any difficulty that the second of the three terms of the right member of (3.5) (after removal of parentheses) is asymptotically equal to n times the last term of the right member of (3.6) and hence that the other two terms of the right member of (3.5) are asymptotically equal to n times the right member of (3.6) without its last term. Now when

$$\frac{i+1}{i}\sqrt{e_1} < 1$$

which will always occur when i is equal to or greater than a sufficiently large fixed integer μ, that part of the right member of (3.6) which is in square brackets is easily seen to be negative. Hence from the definition of asymptotic equivalence it follows that, for all n sufficiently large,

$$\frac{n_2^{(2)}(n_2+1)^{(2)}\,n_1^{(2\mu)}}{n^{(2\mu+2)}} < \frac{(n_2+1)^{(2)}(n_2+1)^{(2)}n_1^{(\mu)}\,n_1^{(\mu)}}{n^{(\mu+1)}\,n^{(\mu+1)}}, \tag{3.20}$$

and

$$\frac{(n_2+1)^{(2)}\,n_1^{(\mu)}}{n^{(\mu+1)}} < 2ne^{\mu+2} < 2ne^{\mu}. \tag{3.21}$$

Hence, for all n sufficiently large,

$$\sigma^2(r_{1\mu}) < 2ne^{\mu}. \tag{3.22}$$

Now consider the expression (3.5) for $i = \mu$ and $i = \mu + 1$. Passage from μ to $\mu + 1$ multiplies the first term of the right member of (3.5) by

$$\frac{(n_1 - 2\mu)(n_1 - 2\mu - 1)}{(n - 2\mu - 2)(n - 2\mu - 3)}, \tag{3.23}$$

and the third term of the right member by

$$\frac{(n_1 - \mu)^2}{(n - \mu - 1)^2}. \tag{3.24}$$

It is easy to see that for large but fixed μ and all n greater than a lower bound which is a function of μ only, the expression (3.23) is less than the expression (3.24). Hence, in view of (3.20), the sum of the first and third terms of the

right member of (3.5) for $i = \mu + 1$ is negative. Now consider what happens to the second term of the right member of (3.5) when i goes from μ to $\mu + 1$. It is multiplied by

$$\frac{(n_1 - \mu)}{(n - \mu - 1)}, \tag{3.25}$$

which, also for large but fixed μ and all n larger than a lower bound which is a function of μ only, is easily seen to be less than e. Consequently

$$\sigma^2(r_{1(\mu+1)}) < 2ne^{\mu+1}. \tag{3.26}$$

It can be seen without difficulty that such a passage of (3.5) to the next higher index is always accompanied by multiplication by expressions similar to (3.23), (3.24), and (3.25), for which similar inequalities hold and that consequently

$$0 \leq \sigma^2(r_{1i}) < 2ne^i, \tag{3.27}$$

and for similar reasons

$$0 \leq \sigma^2(r_{2i}) < 2ne^i,$$

for all i not less than μ and for all n greater than a lower bound which is a function of μ only (although it may be necessary to increase the original μ so that *both* the last two equations hold). The required result follows from (3.19) and the convergence of the series (3.9).

The proof of Theorem 4 follows along the same lines as that of Theorem 2. When $f(x) \equiv 1$, $F(V) \equiv U(V)$, the statistic discussed in [2]. Other such results follow from specialization of $f(x)$. Theorem 4 may also be generalized so that the elements v_i which add up to n_1 are operated on by a function f_1, while the elements v_i which add up to n_2 are operated on by another function f_2, but this is easy to see and we do not go into the details.

4. Tests of hypotheses in the non-parametric case. The great advances that have been made in mathematical statistics in recent years have been in two directions. On the one hand, the foundations of statistics, the theory of estimation and of testing hypotheses have been put on a rigorous basis of probability theory, and on the other, powerful methods for obtaining critical regions and confidence intervals and criteria for appraising their efficacy have been developed. Most of these developments have this feature in common, that the distribution functions of the various stochastic variables which enter into their problems are assumed to be of known functional form, and the theories of estimation and of testing hypotheses are theories of estimation of and of testing hypotheses about, one or more parameters, finite in number, the knowledge of which would completely determine the various distribution functions involved. We shall refer to this situation for brevity as the parametric case, and denote the opposite situation, where the functional forms of the distributions are unknown, as the non-parametric case.

The literature of theoretical statistics, therefore, deals principally with the parametric case. The reasons for this are perhaps partly historic, and partly the fact that interesting results could more readily be expected to follow from the assumption of normality. Another reason is that, while the parametric case was for long developed on an intuitive basis, progress in the non-parametric case requires the use of modern notions. However, the needs of theoretical completeness and of practical research require the development of the theory of the non-parametric case. The purpose of the following section is to contribute to this theory.

Brief mention of some of the literature may be made here. The problem of parametric estimation by confidence intervals, was put on a rigorous foundation by Neyman [4] and extended to the estimation of distribution functions in the non-parametric case by means of confidence belts by Wald and Wolfowitz [5]. Problems of testing non-parametric hypotheses have been treated in various places. The rank correlation coefficient has been used for a long time to test the independence of two variates. Its distribution was shown to be asymptotically normal by Hotelling and Pabst [6] and its small sample distribution was discussed by Olds [7]. The problem of two samples has been discussed, among others, by Thompson [8], Dixon [9] and Wald and Wolfowitz [2]. In 1937, Friedman [10] posed the non-parametric analogue of the problem in the analysis of variance and proposed a very ingenious solution.

All these proposed solutions have this in common, that there exists no general principle which can be applied in each particular case to obtain a critical region, a role which is performed in the parametric case by Fisher's principle of maximum likelihood and the likelihood ratio criterion (Neyman and Pearson, [11]), whose validity, at least for large samples, has been established by Wald ([12], [13]). In each problem the solutions proposed have been intuitive and usually based on an analogy to the corresponding problem in the parametric case. Thus the principal justification for the use of the rank correlation coefficient is that its distribution is independent of the unknown distribution function (under the null hypothesis) and that its structure resembles that of the ordinary correlation coefficient. But any function of the order relations among the variates (cf. [2], p. 148) has a distribution which is independent of the unknown population distribution under the null hypothesis The same objection may be made to papers [8], [9], [10], [2], except that in [2], although the solution there proposed is an intuitive one, the criterion of consistency is extended from the parametric case to the non-parametric one. The fulfilment of this condition is a minimal requirement of a good test and on this basis the solution proposed in one of the previous papers cannot be considered a good one.

In the following section we shall show that the likelihood ratio criterion may be extended to the non-parametric case where the test must be made on the order relations among the observations and that for a certain class of these problems which fulfill the same requirement as that for the application of the likelihood ratio criterion in the parametric case it would thus appear to furnish

a general method by which statistics may be obtained for a specific problem. We shall show this by applying it to the problem of two samples. This will serve to explain the method. Another problem will be discussed later. The ultimate justification of any statistic must be its power function, which ought therefore to constitute the next subject of investigation for these problems. Since for problems in the non-parametric case it is almost certain that uniformly most powerful tests do not exist, the question of determining the alternatives with respect to which proposed tests are powerful is particularly important.

5. The problem of two samples. Let X and Y be two stochastic variables with the distribution functions $f(x)$ and $g(x)$, respectively. (The term distribution function will always denote the cumulative distribution function. The letter P followed by an expression in braces will stand for the probability of the relation in braces. Hence $P\{X < x\} = f(x)$ for all x.) $f(x)$ and $g(x)$ are assumed continuous. The n_1 observations $x_1, x_2 \cdots, x_{n_1}$ and n_2 observations $y_1, y_2, \cdots, y_{n_2}$ are made on X and Y respectively. The (null) hypothesis to be tested is that $f(x) \equiv g(x)$. The admissible alternatives are all continuous distribution functions $f(x)$ and $g(x)$ such that $f(x) \not\equiv g(x)$. The $n_1 + n_2 = n$ observations are arranged in ascending order of size, thus: $Z = z_1, \cdots, z_n$ where $z_1 < z_2 < \cdots < z_n$ (the probability that $z_i = z_{i+1}$ is 0). Let $V = v_1, v_2, \cdots, v_n$ be a sequence defined as follows: $v_i = 0$ if z_i is a member of the set $x_1, x_2, \cdots, x_{n_1}$ and $v_i = 1$ if z_i is a member of the set $y_1, y_2, \cdots, y_{n_2}$. Then any statistic used to test the null hypothesis must be a function only of V ([2], p. 148).

We now apply the method of Neyman and Pearson [11] as follows: Ω is the totality of all couples $(d_1(x), d_2(x))$ of continuous distribution functions. The set ω, a subset of Ω, is the totality of all couples of distribution functions for which $d_1 \equiv d_2$. The sample space is the totality of all sequences V. The null hypothesis states that (f, g) is a member of ω. The admissible alternatives are that (f, g) is a member of Ω not in ω. The distribution of any function of V is the same for all members of ω. Hence this essential requirement on the statistic to be selected for the application of the likelihood ratio criterion (cf. [11]) is satisfied by any statistic which is a function of V alone. Furthermore, all sequences V have the same probability if the null hypothesis is true ([2], p. 149). The numerator of the likelihood ratio is therefore a function only of n_1 and n_2, is the same for all V, and is therefore of no further interest. Hence $T'(V)$, a function of V which is a monotonic function of the likelihood ratio for this problem, may be defined as the denominator of the likelihood ratio, as follows: Let $P\{V; (d_1, d_2)\}$ be the probability of V when $f \equiv d_1$, and $g \equiv d_2$. Then

$$T'(V) = \max_{\Omega} P\{V; (d_1, d_2)\}.$$

The critical values of $T'(V)$ are the large values. However, we may use instead of $T'(V)$ a convenient monotonic function of $T'(V)$.

As an approximation to $T'(V)$ we propose $T(V)$, a statistic which is obtained on the assumption that for a given V a couple (d_1^*, d_2^*) which is essentially the same as that of the two sample distribution functions corresponding to the particular V approximates a couple which maximizes the right member of (5.1). (We say "*a*" couple because it cannot be unique.) This assumption seems a reasonable one, particularly for large samples. Only the form of (d_1^*, d_2^*) is assumed and the missing parameters are obtained in accordance with (5.1). Before describing the matter precisely, it must be stressed that this is offered only as a plausible approximation. For certain extreme V, for example, like those where zeros and ones nearly alternate, this is definitely not the maximizing couple. In spite of this the statistic $T(V)$ assigns to these V values which are furthest removed from the critical region for any level of significance, as indeed any good statistic should.

We first define a "*run*" as in [2], p. 149. A subsequence $v_{(t+1)}, v_{(t+2)}, \cdots, v_{(t+r)}$ of V (where r may also be 1) is called a "*run*" if $v_{(t+1)} = v_{(t+2)} = \cdots = v_{(t+r)}$ and if $v_t \neq v_{(t+1)}$ when $t > 0$ and if $v_{(t+r)} \neq v_{(t+r+1)}$ when $t + r < n$. Let l_{1j} be the number of elements in the j^{th} *run* of elements 0, and l_{2j} the number of elements in the j^{th} *run* of elements 1. Suppose for a moment that the first element in V is a 0. Consider the following situation: There is an interval $[a_1, a_2]$, $a_1 < a_2$, on the line $-\infty < x < +\infty$ such that

$$P\{a_1 \leq X \leq a_2\} > 0, \qquad P\{a_1 \leq Y \leq a_2\} = 0,$$
$$P\{X < a_1\} = P\{Y < a_1\} = 0.$$

This is followed by an interval $[b_1, b_2]$, $b_1 = a_2$, such that $P\{b_1 \leq X \leq b_2\} = 0$, $P\{b_1 \leq Y \leq b_2\} > 0$. This is in turn followed by an interval $[a_3, a_4]$, $a_3 = b_2$, such that $P\{a_3 \leq X \leq a_4\} > 0$, $P\{a_3 \leq Y \leq a_4\} = 0$, etc. It is clear that the lengths and location of the intervals described are immaterial, provided only that they do not overlap. Also the distributions of X and Y within each interval are immaterial, provided only that they are continuous. All that matters for finding $P\{V; (d_1^*, d_2^*)\}$ is that the number and the order of the disjunct intervals shall be the same as those of the runs in V, (i.e., intervals of positive probability for X must alternate with intervals of positive probability for Y, the number of intervals of positive probability for X and for Y must equal respectively the number of runs of the element 0 and the number of runs of the element 1, and the probability of the first interval on the left shall be positive for X or for Y according as the first run in V is of elements 0 or of elements 1, with the same relation obtaining between the last interval on the right and the last run in V) and the probability of these intervals. Let P_{1j} be the sought for probability of the interval which corresponds to the jth run of elements 0 and P_{2j} the probability of the interval which corresponds to the jth run of elements 1. In order to obtain V, it is necessary that the elements constituting each run shall fall into its corresponding interval. Then clearly by the multinomial theorem

$$P\{V; (d_1^*, d_2^*)\} = \prod_i n_i! (\prod_j (l_{ij}!)^{-1} P_{ij}^{l_{ij}}) \tag{5.2}$$

where $i = 1, 2$ and where, when i is fixed, the product with respect to j is taken over all runs of the corresponding element. The right member of (5.2) is to be maximized with respect to the P_{ij}, subject of course to the constraints

$$\sum_j P_{ij} = 1 \qquad (i = 1, 2). \tag{5.3}$$

Then it may easily be verified that the maximum occurs when

$$P_{ij} = \frac{l_{ij}}{n_i} \qquad (i = 1, 2) \tag{5.4}$$

For, after multiplying by a constant and taking the logarithm we introduce two Lagrange multipliers μ_1 and μ_2 so that the maximizing P_{ij} are given by the equations (5.3) and those obtained by equating to zero all the partial derivatives of

$$\sum_i \sum_j (l_{ij} \log P_{ij} - \mu_i P_{ij}).$$

The latter are therefore

$$\mu_i = \frac{l_{ij}}{P_{ij}} \qquad (i = 1, 2),$$

for all j, whence (5.4) follows. It is easy to see that the extremum thus obtained is a maximum and also an absolute maximum. The sought-for statistic $T(V)$ is then the right member of (5.2) after the results (5.4) have been inserted. It may be simplified by removing all factors which are functions only of n_1 and n_2 (since these will then be the same for all V) and recalling that

$$\sum_j l_{ij} = n_i \qquad (i = 1, 2). \tag{5.5}$$

It will be convenient to take the logarithm of the resulting expression, so that with a slight change of notation we finally have

$$T(V) = \sum_i \sum_j \bar{l}_{ij} \tag{5.6}$$

where

$$\bar{l}_{ij} = \log \left(\frac{l_{ij}^{l_{ij}}}{l_{ij}!} \right). \tag{5.7}$$

This result is immediately extensible to the problem of k samples and by way of summary we recapitulate it as follows:

Let there be given k stochastic variables $X_1, \cdots, X_k$ with the respective distribution functions $f_1(x), \cdots, f_k(x)$, about which nothing is known except that they are continuous. Random independent observations, n_i in number, are made on X_i $(i = 1, \cdots, k)$. It is desired to test the hypothesis that $f_1 \equiv f_2 \equiv \cdots \equiv f_k$, the admissible alternatives being all k-tuples of continuous distribution functions. The sequence V is obtained from the sequence Z by

replacing an observation on X_i by the element i. Let l_{ij} be the number of elements in the jth run of elements i. Then the corresponding statistic for testing the null hypothesis is $T_k(V)$ or any monotonic function of it, where

$$T_k(V) = \sum_{i=1}^{k} \sum_{j} \bar{l}_{ij}$$

and $\bar{l}_{ij}$ is given by (5.7). The large values of $T_k(V)$ are the critical values.

Let r_{ij} denote the number of runs of length j in the elements i. Let $\sum_j r_{ij} = r_i$. Of course $\sum_j jr_{ij} = n_i$. Also let $s_j = \sum_i r_{ij}$. Then

$$T_k(V) = \sum_i \sum_j \bar{j} r_{ij} \tag{5.8}$$

and

$$T_k(V) = \sum_j \bar{j} s_j. \tag{5.9}$$

If a table were constructed of the numbers (5.7) from 1 to 50, say, or from 1 to 100, this would cover most of the cases arising in practice. The calculation of $T_k(V)$ by means of (5.9) would then be so simple that it could be performed very expeditiously by an ordinary clerk and with very much less labor than is required for most statistics in common use, like the correlation coefficient, for example. As a matter of interest we note that

$$\bar{1} = 0$$

$$\bar{2} = .693$$

$$\bar{3} = 1.50$$

$$\bar{4} = 2.37$$

$$\bar{5} = 3.26$$

and that

$$\bar{p} < p \tag{5.10}$$

where p is any integer ≥ 1. (5.10) follows from the fact that

$$p! > (\sqrt{2\pi p} - 1)p^p e^{-p}.$$

The distribution of $T(V)$ may be found for small samples by enumerating the sequences V, all of which have the same probability under the null hypothesis, and assigning to each V its $T(V)$. The critical region consists of the V's for which $T(V)$ takes the largest values, taken in sufficient number to make the critical region of proper size. It will not be necessary to enumerate all the V's, since it is readily apparent that certain V's can never belong to a critical region of any reasonable size. (Roughly speaking, a V with a large number of runs of short length will yield a small $T(V)$ and vice versa.) For large samples, the result of Section 3 is available, with $f(x) = \bar{x}$. From (5.10) it follows

easily that the corresponding series (3.9) is convergent, so that $\overline{T}(V)$ is asymptotically normally distributed. It must be remembered when using tables of the normal distribution that the critical region of $\overline{T}(V)$ lies in only one "tail" of the normal curve. The greatest difficulty will occur for samples of moderate size. Methods like those of Olds [7] will probably help there. It is highly unlikely that any practicable formula which would give the exact distribution of $T(V)$ exists.

A few brief remarks may be made here on a related problem. Suppose we have observations from two bivariate populations about the distributions of both of which nothing is known except that they are continuous and it is sought to test whether the two populations have the same distribution functions. Suppose further that it were required that the statistic used for this purpose be invariant under any topologic transformation of the whole plane into itself. At this point we quote the following topologic theorem, the proof of which was communicated to the author by Dr. Herbert Robbins: *Let x_1, y_1, x_2, y_2, $\cdots$, x_p, y_p be any $2p$ distinct points in the plane. There exists a topologic transformation of the whole plane into itself which takes x_i into y_i $(i = 1, 2, \cdots, p)$.* As a consequence of this theorem we get the absurd result that the required statistic must be a constant. Hence this statistical problem can have no solution.

As a matter of interest this statistical problem would have no solution even if it were not for the topologic theorem. The fact is that a continuous distribution on a line remains continuous under a topologic transformation of the whole line into itself, but a continuous distribution in a k-dimensional (Euclidean) space ($k > 1$) may become discontinuous under a topologic transformation of the whole space into itself. (The probability distribution in the first space always determines a probability distribution in the transformed space, for probability functions are defined over all Borel sets of the space (cf. [15], p. 7) and a topologic transformation carries Borel sets into Borel sets (cf. [16], p. 195, Theorem II)). Consider the following example in the plane: A bivariate distribution function assigns probability 1 to a line L oblique to the coordinate axes, while any interval which contains no segment of the line L has probability 0. On the line L the (one-dimensional) probability distribution may be arbitrary, provided it is continuous. The bivariate distribution function is without difficulty seen to be continuous. Now rotate the coördinate axes until one of them is parallel to L. It is easy to see that after the rotation the bivariate distribution function is discontinuous.

The question of whether a useful statistical problem could be obtained by properly delimiting the class of transformations which are to leave the statistic invariant and the solution of such a problem remain to be investigated.

6. The problem of the independence of several variates. This is an important practical problem and one of the earliest discussed in the literature (cf., for example, [6]). Let X_1 and X_2 be stochastic variables with the joint (cumulative) distribution function $F(x_1, x_2)$ which is known to be continuous in both variables

jointly (i.e., $F(x_1, x_2) = P\{X_1 < x_1 ; X_2 < x_2\}$, where the right member is the probability of the occurrence of *both* the relations in braces). The marginal distributions $f_1(x_1)$ and $f_2(x_2)$ of X_1 and X_2 respectively are defined as follows:

$$f_1(x_1) = P\{X_1 < x_1\} = \lim_{x_2 \to +\infty} F(x_1, x_2),$$

$$f_2(x_2) = P\{X_2 < x_2\} = \lim_{x_1 \to +\infty} F(x_1, x_2).$$

(It is easy to see that the continuity of $F(x_1, x_2)$ implies the continuity of $f_1(x_1)$ and $f_2(x_2)$.)

The n random, independent pairs of observations $x_{11}, x_{21}, \cdots x_{1n}, x_{2n}$ are made on X_1 and X_2. The null hypothesis states that

$$F(x_1, x_2) \equiv f_1(x_1) \cdot f_2(x_2) \tag{6.1}$$

i.e., that X_1 and X_2 are independent. The alternative hypotheses are that $F(x_1, x_2)$ does not satisfy (6.1).[1]

Let the set $x_{11}, x_{12}, x_{13}, \cdots, x_{1n}$ be arranged in order of ascending size, thus: $Z = z_1, z_2, z_3, \cdots, z_n$ where $z_1 < z_2 < \cdots < z_n$. The jth member of this sequence will be said to have the rank j. In the same manner ranks are assigned to the x_{2j} $(j = 1, \cdots, n)$. (It is easy to see that, since $f_1(x_1)$ and $f_2(x_2)$ are continuous, the probability that $z_j = z_{j+1}$ is 0 etc.) In the sequence Z the element z_j $(j = 1, \cdots, n)$ is replaced by the rank of its associated observation on X_2. We obtain a permutation of the integers $1, 2, \cdots, n$ which we denote by R. If in the procedure for obtaining R, we had reversed the roles of the x_{1j} and x_{2j}, we would have obtained the permutation R'. It is easy to see that any statistic, say M'', used to test the null hypothesis, must be a function only of R, with the added proviso that $M''(R) = M''(R')$. (The rank correlation coefficient is such a statistic.) Under the null hypothesis all the R have the same probability $\left(= \dfrac{1}{n!}\right)$.

The procedure of applying the likelihood ratio principle to this problem would then be as follows: Ω is the totality of all bivariate distribution functions $H(x_1, x_2)$ which are continuous in both variables jointly. The respective marginal distributions corresponding to $H(x_1, x_2)$ will be denotedby $h_1(x_1)$ and $h_2(x_2)$. ω is a subset of Ω which consists of all $H(x_1, x_2)$ for which $H(x_1, x_2) \equiv h_1(x_1) \cdot h_2(x_2)$. The sample space is the totality of all sequences R. The null hypothesis states that $F(x_1, x_2)$ is a member of ω. The admissible alternatives are that $F(x_1, x_2)$ is a member of Ω not in ω. The distribution of any function of R is the same for all members of ω. Thus the essential requirement for the applicability of the likelihood ratio criterion is fulfilled. All sequences R have the same probability for all members of ω; hence the numerator of the likelihood ratio is a func-

[1] It is easy to see that the independence or dependence of two stochastic variables is not a property which will remain invariant under a topologic transformation of the plane into itself. We therefore require of the statistic only that it be invariant under topologic transformation of *each* variable into itself, separately.

tion only of n which may therefore be ignored. We may then define $M'(R)$, a monotonic function of the likelihood ratio as the denominator of the likelihood ratio, thus:

$$M'(R) = \max_{\Omega} P\{R; H(x_1, x_2)\} \tag{6.2}$$

where $P\{R; H(x_1, x_2)\}$ is the probability of R when $H(x_1, x_2)$ is the joint distribution function of X_1 and X_2. The critical values of $M'(R)$ are the large values.

We now propose an approximation to $M'(R)$ which we shall call $M(R)$. We do this by describing a distribution function $H^*(x_1, x_2)$ for each R which seems a plausible approximation to a maximizing distribution function. It may be derived from certain assumptions about the nature of the maximizing distribution function which we omit. The remarks made in the preceding section about the character of the approximation apply here as well. As before we specify only the form of the function and leave certain parameters, finite in number, to be determined in accordance with (6.2). (If the construction of $H^*(x_1, x_2)$ should appear somewhat involved, this is due only to the analytic description. A sketch will show the essential simplicity of the situation.) We then have

$$M(R) = P\{R; H^*(x_1, x_2)\}.$$

Let $R = a_1, a_2, \cdots, a_n$ be a given permutation of the integers 1 to n. A sub-sequence $a_{(i+1)}, a_{(i+2)}, \cdots, a_{(i+l)}$ will be called a run of length l if the following conditions are fulfilled:

(6.3) The indices of the a's are consecutive,

(6.4) If l' is any integer such that $1 \leq l' < l$, then

$$|a_{(i+l')} - a_{(i+l'+1)}| = 1,$$

(6.5) if $i > 0$, $|a_i - a_{(i+1)}| > 1$,

(6.6) if $i + l < n$, $|a_{(i+l)} - a_{(i+l+1)}| > 1$.

The run will be called an ascending run or a descending run according as $a_{(i+1)} - a_{(i+2)} = -1$ or $+1$. A run of length 1 is of either type, at pleasure. For example, let

$$R = 5, 6, 1, 4, 3, 2.$$

The first run is 5, 6, the second 1, the last 4, 3, 2. 5, 6 is an ascending run of length two, 4, 3, 2 a descending run of length three, and 1 a run of length one.

$H^*(x_1, x_2)$ is a degenerate distribution function such that the relation between X_1 and X_2 is functional (this is a special case of stochastic relationship). That is to say, $X_2 = \varphi(X_1)$, where $\varphi(X_1)$ is a single-valued function defined for all the possible values of X_1, with a single-valued inverse $\varphi^{-1}(X_2)$ defined for all possible values of X_2. Hence $H^*(x_1, x_2)$ is completely specified when the function $X_2 = \varphi(X_1)$ and $h_1^*(x_1)$ the marginal distribution function of X_1, are given ($h_1^*(x_1)$ must of course be continuous).

Consider a system of intervals on the line $-\infty < x_1 < +\infty$ of which $(i - 1, i)$

is the ith, $i = 1, 2, \cdots, n$ and a similar system on the line $-\infty < x_2 < +\infty$. (Actually, as in the previous section, neither the length of the intervals nor their location is material. The intervals need merely be disjunct and in a certain order. We are using these particular intervals to simplify the notation.) Let l_1 be the length of the first run. a_1 is its first element. Then let

$$p_1 = P\{0 \leq X_1 \leq l_1 ; h_1^*(x_1)\}$$

be one of the as yet undetermined parameters. We now partly define $h_1^*(x_1)$ as follows:

$$\begin{aligned} h_1^*(x_1) &= 0, & x_1 &\leq 0 \\ h_1^*(x_1) &= 1, & x_1 &\geq n \\ h_1^*(l_1) &= p_1. \end{aligned} \tag{6.7}$$

Within the interval $(0, l_1)$, $h_1^*(x_1)$ may be any continuous monotonic increasing function which satisfies (6.7). We partly define $\varphi(X_1)$ as follows:
If the first run is ascending, let

$$\varphi(0) = a_1 - 1 \tag{6.8}$$

$$\varphi(x_1) = a_1 - 1 + x_1, \qquad 0 < x_1 \leq l_1. \tag{6.9}$$

If the first run is descending, let

$$\varphi(0) = a_1 \tag{6.10}$$

$$\varphi(x_1) = a_1 - x_1, \qquad 0 < x_1 \leq l_1. \tag{6.11}$$

We proceed in this manner through all the runs of R. Let l_i be the length of the ith run. Let $\lambda_j = \sum_{i<j} l_i$. The first element of the jth run is $a_{(\lambda_j+1)}$. Let

$$p_j = P\{\lambda_j < X_1 \leq \lambda_j + l_j ; h_1^*(x_1)\},$$

be another of the as yet undetermined parameters. We then define $h_1^*(x_1)$ as follows:

$$h_1^*(\lambda_j) = \sum_{i<j} p_i \tag{6.12}$$

$$h_1^*(\lambda_j + l_j) = \sum_{i \leq j} p_i. \tag{6.13}$$

Within the interval $(\lambda_j, \lambda_j + l_j)$, $h_1^*(x_1)$ may be any continuous monotonic increasing function which satisfies (6.12) and (6.13). We define $\varphi(X_1)$ as follows:
If the jth run is ascending, let

$$\varphi(x_1) = a_{(\lambda_j+1)} - 1 + x_1 \qquad (\lambda_j < x_1 \leq \lambda_j + l_j). \tag{6.14}$$

If the jth run is descending, let

$$\varphi(x_1) = a_{(\lambda_j+1)} - x_1 \qquad (\lambda_j < x_1 \leq \lambda_j + l_j). \tag{6.15}$$

If $l_j = 1$, the run may be considered ascending or descending at pleasure.

In order to obtain R, it is necessary that all the elements of a run shall fall into its corresponding interval. Then it is easy to see that by the multinomial theorem

$$P\{R; H^*(x_1, x_2)\} = n! \prod_i (l_i!)^{-1} p_i^{l_i}. \tag{6.16}$$

The right member of (6.16) is to be maximized with respect to the p_i subject to the constraint

$$\Sigma p_i = 1. \tag{6.17}$$

It is easy to verify that the maximum occurs when

$$p_i = \frac{l_i}{n}. \tag{6.18}$$

$M(R)$ is the right member of (6.16) after the results (6.18) have been inserted. It is convenient to remove all factors which are functions only of n and to take the logarithm of the resulting expression. Then with a slight change of notation we may say that

$$M(R) = \sum_i \bar{l}_i \tag{6.19}$$

where

$$\bar{l}_i = \log\left(\frac{l_i^{l_i}}{l_i!}\right). \tag{6.20}$$

The critical values of $M(R)$ are the large values. One may verify without much difficulty that $M(R) = M(R')$, i.e., that the statistic is symmetric with respect to X_1 and X_2 as indeed it should be.

This result is immediately extensible to the problem of testing whether k stochastic variables $X_1, \cdots, X_k$ are independent. We shall not go into the details, which are similar to those described above, and content ourselves with giving the definition of a run for the case $k = 3$. After the observations on X_1 have been arranged in ascending order, we obtain two sequences R_2 and R_3, the associated ranks of the observations on X_2 and X_3. Let $R_2 = b_1, b_2, \cdots, b_n$ and $R_3 = b_1', b_2', \cdots, b_n'$. The ascending sequence of consecutive integers $(i + 1), (i + 2), \cdots, (i + l)$ determines a run of length l if the sequences $b_{(i+1)}, b_{(i+2)}, \cdots, b_{(i+l)}$ and $b'_{(i+1)}, b'_{(i+2)}, \cdots, b'_{(i+l)}$ *both* satisfy (6.4), and if at least one of the sequences satisfies (6.5), and at least one, but not necessarily the same one, satisfies (6.6). The adjectives ascending and descending apply to each sequence separately.

Let r_j be the number of runs of length j in R. Then

$$M(R) = \sum_j jr_j. \tag{6.21}$$

Most of the remarks made in Section 5 about the small sample distribution of $T(V)$ are also applicable to the distribution of $M(R)$. More will be said in the

next section about the distribution of $M(R)$ which involves the solution of a combinatorial problem not discussed in the literature.

7. On the distribution of $W(R)$. While most of the remarks made about the small sample distribution of $T(V)$ apply to the question of the distribution of $M(R)$ in small samples, the situation with respect to the distribution of $M(R)$ in samples of medium size and large size is very different and, in certain respects, is more favorable for practical application than is the case with $T(V)$. It would be reasonable to expect, for example, in view of Section 3 and of the structure of the statistic $M(R)$ that the asymptotic distribution of $M(R)$ should be normal. Surprisingly enough, this is not the case. It is not even continuous. In order to clarify the situation, we begin with a few necessary ideas and definitions.

Let the stochastic variable $W(R)$ be defined as the total number, in R, of runs of the sense of Section 6. We shall be interested in the distribution of $W(R)$. The number n of the pairs of observations on X_1 and X_2 (we consider the case of two variates) will be assumed arbitrary but fixed throughout the discussion and will not be exhibited. Let $N(k)$ be the number of sequences R (of the integers 1 to n) which contain exactly k runs.

Consider, for example, for the case $n = 6$, the sequence 2 3 4 6 5 1. We shall say that this sequence contains the "contacts" (2, 3), (3, 4), (6, 5). In general, a contact is defined as the juxtaposition, in the sequence R, of consecutive numbers, whether in ascending or descending order. If k is the number of runs and l the number of contacts in a sequence R, then obviously

$$k + l = n. \tag{7.1}$$

Let R_0 be the sequence $1, 2, \cdots, n$ of the first n integers in ascending order. The $n - 1$ contacts of this sequence may themselves be arranged in a sequence R^* of contacts, thus:

$$(1, 2), (2, 3), \cdots, (n - 1, n).$$

Suppose l of the contacts which constitute the sequence R^* are selected in some manner to form the set O. The remaining $n - 1 - l$ contacts form the complementary set O'. After this selection the sequence R^* may be considered a sequence of the type of the sequences V of Section 5 with the members of O playing the role of the elements 0 and the members of O' playing the role of the elements 1. When R^* is considered in this manner we will write it as $R^*(O)$. The definition of a run of Section 5 as applied to sequences V is now applicable to $R^*(O)$. We will call any such run of the members of O or of O' a group.

We wish first to answer the following question: In how many ways can the set O be selected from among the elements of R^* so that it will contain l members arranged in $R^*(O)$ in i groups? If, for a given O, i' be the number of groups into which O' is divided in $R^*(O)$, it is clear that $i - i'$ can equal only -1, 0, or $+1$. Hence only four situations can arise, as follows:

a) $i' = i + 1$. The first group in $R^*(O)$ is therefore composed of elements of

O'. The number of ways in which l elements can be divided into i runs of the type of Section 2 is the coefficient of x^l in the purely formal expansion of

$$(x + x^2 + x^3 + \cdots)^i = \left(\frac{x}{1-x}\right)^i$$

and is therefore $\binom{l-1}{i-1}$. Similarly $n - 1 - l$ elements can be divided into $i' = i + 1$ runs in $\binom{n-l-2}{i}$ ways. Hence this situation will arise in $\binom{l-1}{i-1}\binom{n-l-2}{i}$ ways.

b) $i' = i - 1$. By a similar argument as above, this can occur in $\binom{l-1}{i-1}\binom{n-l-2}{i-2}$ ways.

c) $i' = i$ and the first group is made up of elements from O. This will occur in $\binom{l-1}{i-1}\binom{n-l-2}{i-1}$ ways.

d) $i' = i$ and the first group is made up of elements from O'. This will also occur in $\binom{l-1}{i-1}\binom{n-l-2}{i-1}$ ways.

The set O which contains l elements arranged in i groups can therefore be selected in

$$\binom{l-1}{i-1}\left(\binom{n-l-2}{i} + \binom{n-l-2}{i-2} + 2\binom{n-l-2}{i-1}\right) \tag{7.2}$$

ways, and the quantity (7.2) is, by elementary combinatorics, equal to

$$\binom{l-1}{i-1}\binom{n-l}{i}. \tag{7.3}$$

Let any set O of l contacts divided into i groups be selected from R^*. Imagine that each contact in O sets up, in R_0, an unbreakable bond which links the two elements involved in the contact, but no contact in O' creates such a bond. Given these bonds set up by O, we seek the number of different sequences into which the n elements of R_0 can be permuted while respecting these bonds. Since there are l bonds, we can actually manipulate only $n - l$ entities, except that two elements linked by a bond may have their order reversed; for example, if O contains (1, 2), 1 may either precede or follow 2 and the bond would still be respected. However, if one contact in a group is reversed, the group as a whole must be reversed, else a bond would be broken. Hence the number of distinct sequences into which R_0 may be permuted while all the bonds set up by O are respected is $2^i(n - l)!$.

Let us refer to the sequences thus obtained as the family generated by O. All the sequences in a family are distinct. Now let O range over all sets of l

contacts selected from R^*. The various families obtained will not be disjunct, but some will have sequences in common. In spite of this, we seek the total of the number of sequences in all the families. The total of the number of sequences in all the families generated by sets of l contacts divided into i groups is, by (7.3) and the result of the preceding paragraph,

$$2^i \cdot \binom{l-1}{i-1}\binom{n-l}{i}(n-l)!. \tag{7.4}$$

Sets of l contacts may consist of 1, 2, $\cdots$ l groups, so that the total number of sequences in all the families generated by sets of l contacts is

$$A_l = \sum_{i=1}^{l} 2^i \binom{l-1}{i-1}\binom{n-l}{i}(n-l)! \tag{7.5}$$

where l may take the values $1, 2, \cdots, (n-1)$. The conventions on the combinatorial symbols will be:

$$\binom{a}{0} = 1, \qquad a \geq 0,$$

$$\binom{a}{b} = 0, \qquad a < b.$$

Define A_0 as

$$A_0 = n!. \tag{7.6}$$

The following equation is trivial:

$$A_0 = \sum_{i=1}^{n} N(i). \tag{7.7}$$

We now consider all the families generated by sets O which contain exactly l contacts. As was said before, the total of the number of sequences in each is A_l. Let $H(l)$ be the set of all the sequences in all these families, with each sequence in $H(l)$ counted as many times as the number of families in which it occurs. Every sequence in $H(l)$ has the l contacts of the set O which generated it, but after permuting R_0 other contacts may still exist. Hence every sequence in $H(l)$ has at least l contacts and therefore by (7.1), at most $n - l$ runs. Clearly, a sequence which has exactly l contacts occurs exactly once in $H(l)$, since it can appear only in the family generated by the set O of its l contacts and in no other family. A sequence which has exactly $(l + 1)$ contacts will appear exactly $\binom{l+1}{l}$ times in $H(l)$, for it will appear once in each family generated by a set O which consists of one of the $\binom{l+1}{l}$ selections of l contacts from among its $(l + 1)$ contacts, and in no other family. Similarly each sequence which has exactly $(l + 2)$ contacts will appear in $H(l)$ $\binom{l+2}{l}$ times, and so forth. We therefore obtain, in view of (7.1),

$$A_l = \sum_{i=l}^{n-1} \binom{i}{l} N(n-i) \qquad (l = 1, 2, \cdots, (n-1)). \tag{7.8}$$

The system of n linear equations (7.7) and (7.8) completely determines the quantities $N(1), N(2), \cdots, N(n)$. The matrix of these equations has a determinant whose absolute value is 1, so that the quantities $N(1), N(2), \cdots, N(n)$ may readily be expressed in determinantal form. Furthermore the moments of $W(R)$ are readily found from these equations. Thus from (7.8) for $l = 1$ we find

$$E(W(R)) = \frac{n^2 - 2n + 2}{n} \sim n - 2 \tag{7.9}$$

and from (7.8) for $l = 2$ and $l = 1$ we find, after a little obvious manipulation,

$$\sigma^2(W(R)) = \frac{2n^3 - 8n^2 + 6n + 4}{n^3 - n^2} \sim 2. \tag{7.10}$$

Higher moments of $W(R)$ may be found in similar manner.

Since the limiting variance of $W(R)$ is 2 it follows that the asymptotic distribution is not continuous. For n of any size the bulk of the values are concentrated in a short interval ending at n. When $W(R) = n$, $M(R) = 0$, when $W(R) = n - 1$, $M(R) = \log 2$, and when $W(R) = n - 2$, $M(R) = \log 4\frac{1}{2}$ or $\log 4$. It is easy to see that for the values of $W(R)$ which differ very little from n there are only a small number of values of $M(R)$, whose asymptotic distribution is also discontinuous. The statistic $W(R)$ is therefore a good approximation to the statistic $M(R)$ for the purposes of tests of significance (for $M(R)$ the large values are the critical values and for $W(R)$ the small values are critical), and has a few additional practical advantages. It is even easier to compute than $M(R)$; the computation is best performed by counting contacts. Since the limiting variance is a small constant, it follows that many tests of significance can be performed simply by use of Tchebycheff's inequality. For example, suppose a given large sample contains 9 contacts, i.e., $n - 9$ runs (we say a "large" sample in order to use the simple limiting mean and variance; if desired or for a small sample these latter may be computed exactly by (7.9) and (7.10)). Then by Tchebycheff's inequality it follows that the probability of obtaining $n - 9$ or fewer runs is less than .041. Thus the presence of 9 contacts would be sufficient to render a sample of great size significant on a 5% level. For the few numbers of contacts about which doubt will exist as to whether or not they are critical values two procedures are possible. Either the equations (7.7) and (7.8) may be solved exactly for the doubtful values, or several higher moments may be found from (7.8) and the methods of Wald [14] can be applied to delimit the missing probabilities to any accuracy desired. By enumerating the few values of $M(R)$ which correspond to several of the largest values of $W(R)$ the distribution of $M(R)$ may be computed sufficiently to serve the purposes of tests of significance.

REFERENCES

[1] Maurice Frechet, *Généralités sur les Probabilités, Variables Aleatoires*, Paris (1937).
[2] A. Wald and J. Wolfowitz, *Annals of Math. Stat.*, Vol. 11 (1940), p. 147.
[3] A. M. Mood, *Annals of Math. Stat.*, Vol. 11 (1940), p. 367.
[4] J. Neyman, *Phil. Trans. Roy. Soc. London*, Vol. 231 (1937), pp. 333–380.
[5] A. Wald and J. Wolfowitz, *Annals of Math. Stat.*, Vol. 10 (June, 1939), p. 105.
[6] H. Hotelling and M. Pabst, *Annals of Math. Stat.*, Vol. 7 (1936), p. 29.
[7] E. G. Olds, *Annals of Math. Stat.*, Vol. 9 (1938), p. 133.
[8] W. R. Thompson, *Annals of Math. Stat.*, Vol. 9 (1938), p. 281.
[9] W. J. Dixon, *Annals of Math. Stat.*, Vol. 11 (June, 1940), p. 199.
[10] Milton Friedman, *Jour. Amer. Stat. Assoc.*, Vol. 32 (1937), p. 675.
[11] J. Neyman and E. Pearson, *Trans. Royal Soc.*, A., Vol. 231 (1933), p. 295.
[12] A. Wald, *Bull. Amer. Math. Soc.*, Vol. 46 (1940), p. 235.
[13] A. Wald, *Bull. Amer. Math. Soc.*, Vol. 47 (1941), p. 396.
[14] A. Wald, *Trans. Amer. Math. Soc.*, Vol. 46 (1939), p. 280.
[15] Harald Cramer, *Random Variables and Probability Distributions*, Cambridge (1937).
[16] F. Hausdorff, *Mengenlehre* (Second Edition), Berlin and Leipzig, 1927.

ON THE THEORY OF TESTING COMPOSITE HYPOTHESES WITH ONE CONSTRAINT

By Henry Scheffé

Princeton University

1. Introduction. Our purpose is to extend some of the Neyman-Pearson theory of testing hypotheses to cover certain cases of frequent interest which are complicated by the presence of nuisance parameters. Our results give methods of finding critical regions of types B and B_1. Type B regions were defined by Neyman [1] for the case of one nuisance parameter. Type B_1 regions are the natural generalization of the type A_1 regions of Neyman and Pearson [5] to permit the occurrence of nuisance parameters. The reader familiar with the work of these authors will recognize most of the notation and some of the methods.

We consider a joint distribution of n random variables $x_1, x_2, \cdots, x_n$, depending on l parameters $\theta_1, \theta_2, \cdots, \theta_l$, $l \leqq n$. The functional form of the distribution is given. The random variables may be regarded as the coordinates of a point E in an n-dimensional sample space W, the parameters, as the coordinates of a point Θ in an l-dimensional space Ω of admissible parameter values. Ω, unlike W, in general will not be a complete Euclidian space. Let ω denote the subspace of Ω defined by $\theta_1 = \theta_1^0$. The hypothesis we consider is

$$H_0 : \Theta \epsilon \omega.$$

Neyman and Pearson [4] call H_0 a hypothesis with $l - 1$ degrees of freedom; for our present purpose we shift the emphasis by saying it has one constraint.

It is clear that whenever we test whether a parameter has a given value, and other parameters occur in the distribution, we are testing a hypothesis with one constraint. Hypotheses of the type $\theta_1 = \theta_2$, in which we do not specify the common value of θ_1 and θ_2, nor the values of any other parameters, may always be transformed to H_0 by choosing new parameters. In general, the hypothesis that the parameter point Θ lies on some hypersurface in Ω, $g(\theta_1, \theta_2, \cdots, \theta_l) = g_0$, may be transformed to H_0 if the function g satisfies certain conditions,—say, g is continuous and monotone-increasing in one of the θ's for all Θ in Ω. Another circumstance lending importance to the theory of testing hypotheses with one constraint is its connection with the theory of confidence intervals, which we shall point out below.

The path which led Neyman to critical regions of type B is the following: Every Borel-measurable region w of sample space determines a test of H_0, which consists of rejecting H_0 if and only if E falls in w. In deciding which is a most efficient test, one may limit the competition to similar[1] regions, if such exist. Because of the general non-existence [2, p. 372] of uniformly most

[1] Defined by condition (a) of definition 1.

powerful tests, one is led to consider common best critical regions [4] if he is interested only in alternatives $\theta_1 < \theta_1^0$ (or $\theta_1 > \theta_1^0$), or else regions giving an unbiased test [1, p. 251]. Narrowing the competition further to the latter class of regions, one is led to regions of type B if he seeks tests which are most powerful for θ_1 very near to θ_1^0, and to type B_1 regions if he is not content with this. These types of regions are defined in section 2.

We may now state the relationship of hypotheses with one constraint to the theory of confidence intervals [2]. To find confidence intervals for θ_1, we must first find similar regions $w(\theta_1^0)$ for testing H_0. If with every admissible θ_1 we can associate a $w(\theta_1)$, then confidence regions for θ_1 are determined, and if these be intervals, they are confidence intervals. Every class of similar regions mentioned above is intimately related to a category of confidence intervals. In particular, to find Neyman's short unbiased confidence intervals we must first solve the problem of type B regions. Likewise, if we define shortest unbiased confidence intervals in the obvious way along the lines laid down by Neyman, their discovery rests on the solution of the problem of type B_1 regions.

While the assumptions of section 3, especially 3^0, are unpleasantly restrictive—they are obviously tailored to fit the proof rather than the problem—they are nevertheless satisfied in many sampling problems associated with normal distributions. An application of the theorems of section 4 will be given in another paper *On the ratio of the variances of two normal populations*. The present theory was needed to round out that paper and was originally planned as a section thereof. However, it seems desirable for the convenience of other workers who might have use for the theory not to bury it under the preceding title.

Section 5 consists of an appendix on the moment problem raised by assumption 5^0.

2. Definitions. The symbols w, w_0, w_1 will always be understood to denote Borel-measurable regions in W. We shall symbolize $\partial^i Pr\{E \epsilon w \mid \Theta\}/\partial\theta_1^i$ for $i = 0, 1, 2$ by $P(w \mid \Theta)$, $P'(w \mid \Theta)$, $P''(w \mid \Theta)$, respectively. Since θ_1 plays a distinguished rôle, it will often be convenient to write $\Theta = (\theta_1, \vartheta)$, where the nuisance parameters are denoted by $\vartheta = (\theta_2, \theta_3, \cdots, \theta_l)$.

Definition 1: *w_0 is said to be a type B region for testing H_0 if for all Θ in ω*

(a) $P(w_0 \mid \theta_1^0, \vartheta) = \alpha$, *where α is independent of ϑ,*

(b) $P'(w_0 \mid \theta_1^0, \vartheta)$, $P''(w_0 \mid \theta_1^0, \vartheta)$ *exist,*

(c) $P'(w_0 \mid \theta_1^0, \vartheta) = 0$,

(d) $P''(w_0 \mid \theta_1^0, \vartheta) \geqq P''(w_1 \mid \theta_1^0, \vartheta)$ *for all w_1 satisfying* (a), (b), (c).

Definition 2: *w_0 is said to be of type B_1 if the conditions* (a), (b′), (c), (d′) *are satisfied. The conditions* (a), (c) *are given in definition* 1, *the other two are*

(b′) $P'(w_0 \mid \theta_1, \vartheta)$ *is continuous in θ_1 at $\theta_1 = \theta_1^0$ for all Θ in ω,*

(d′) $P(w_0 \mid \theta_1, \vartheta) \geqq P(w_1 \mid \theta_1, \vartheta)$ *for all w_1 satisfying* (a), (b′), (c), *and all Θ in Ω.*

3. Assumptions. $p(z_1, z_2, \cdots, z_m \mid \Theta)$ will be a generic notation for the p.d.f. (probability density function) of random variables $z_1, z_2, \cdots, z_m$ whose distribution depends on Θ. The numbering of the following assumptions follows that of Neyman elsewhere [1].

1^0. (a) There exists a p.d.f. $p(E \mid \Theta)$ such that for any w, and any $\Theta \epsilon \Omega$,

$$P(w \mid \Theta) = \int_w p(E \mid \Theta) dW \tag{1}$$

where dW denotes the volume element $dx_1 dx_2 \cdots dx_n$.

(b) The region W_+ in W defined by $p(E \mid \Theta) > 0$ is independent of Θ for $\Theta \epsilon \omega$.

(c) The connectivity of ω is such that it is possible to pass from any point Θ' in ω to any other point Θ'' in ω by a path lying entirely in ω and consisting of a finite number of segments on each of which all but one of $\theta_2, \theta_3, \cdots, \theta_l$ are constant.

2^0. For all $E \epsilon W_+$ and $\Theta \epsilon \omega$, $p(E \mid \Theta)$ is differentiable twice with respect to θ_1 and indefinitely with respect to $\theta_2, \theta_3, \cdots, \theta_l$. For any w, and any $\Theta \epsilon \omega$, the corresponding derivatives of $P(w \mid \Theta)$ exist and may be obtained by differentiating under the integral sign in (1).

We now define

$$\phi_i = \partial \log p(E \mid \Theta)/\partial\theta_i, \qquad \phi_{ij} = \partial\phi_i/\partial\theta_j, \quad i, j = 1, 2, \cdots, l.$$

3^0. For all $E \epsilon W_+$ and $\Theta \epsilon \omega$, $\phi_i = \phi_i(E, \Theta)$ is continuous in E, $i = 1, 2, \cdots, l$, and

$$\phi_{ij} = A_{ij} + \sum_{k=2}^{l} B_{ijk}\phi_k, \qquad i, j = 2, 3, \cdots, l, \tag{2}$$

$$\phi_{i1} = A_{i1} + \sum_{k=1}^{l} B_{i1k}\phi_k, \qquad i = 1, 2, \cdots, l, \tag{3}$$

where $A_{ij} = A_{ij}(\theta_1^0, \vartheta)$, $B_{ijk} = B_{ijk}(\theta_1^0, \vartheta)$ are continuous in each of $\theta_2, \theta_3, \cdots, \theta_l$.

4^0. The matrix $(\partial\phi_i/\partial x_j)$, $i = 1, 2, \cdots, l$; $j = 1, 2, \cdots, n$, contains an $l \times l$ minor which is non-singular[2] for all $E \epsilon W_+$ and $\Theta \epsilon \omega$, and whose elements are continuous in E.

Write $\Phi = (\phi_2, \phi_3, \cdots, \phi_l)$, and denote by $p(\phi_1, \Phi \mid w, \Theta)$ the p.d.f. of (ϕ_1, Φ) calculated under the assumption that $E \epsilon w$, i.e., that the p.d.f. of E is $p(E \mid \Theta)/P(w \mid \Theta)$ for $E \epsilon w$ and zero for $E \epsilon W - w$. Define

[2] If for each $\Theta \epsilon \omega$, 4^0 is violated on an exceptional set $U(\Theta)$ for which $P(U(\Theta) \mid \Theta) = 0$, the theorems 1 and 2 may still be valid. What is essential is the existence of the p.d.f. $p(\phi_1, \phi_2, \cdots, \phi_l \mid \Theta)$ for all $\Theta \epsilon \omega$. On reconsidering the theorems and their proofs, the reader will see that if the set $U(\Theta)$ is deleted from W_+, then $1^0(b)$ may be violated, but not seriously, and no essential changes are necessary. The addition of the necessary qualifying clauses to our statements, regarding sets of probability zero, would encumber the developments.

$$Q_s(\Phi \mid w, \Theta) = \int_{-\infty}^{+\infty} \phi_1^s\, p(\phi_1, \Phi \mid w, \Theta)\, d\phi_1. \tag{4}$$

Let w_1 be any region satisfying condition (a) of definition 1:

5^0. We assume, for each $\Theta \in \omega$, that if the moments[3] of $Q_s(\Phi \mid w_1, \Theta)$ and $Q_s(\Phi \mid W, \Theta)$ are the same then these functions are equal for almost all Φ

(a) for $s = 0$,

(b) for $s = 1$.

Note that Q_0 is p.d.f., Q_1 is not.

4. Theorems. A result of Neyman's [1] for $l = 2$ is generalized in the following[4]

THEOREM 1: *Under the assumptions 1^0 to 5^0, consider the existence of functions $k_i(\Phi, \theta_1^0, \vartheta)$, $i = 1, 2$, such that $k_1 < k_2$ and*

$$\int_{k_1(\Phi,\theta_1^0,\vartheta)}^{k_2(\Phi,\theta_1^0,\vartheta)} \phi_1^s\, p(\phi_1, \Phi \mid \theta_1^0, \vartheta)\, d\phi_1 = (1 - \alpha) \int_{-\infty}^{+\infty} \phi_1^s\, p(\phi_1, \Phi \mid \theta_1^0, \vartheta)\, d\phi_1, \qquad s = 0, 1, \tag{5}$$

for all $\Phi = (\phi_2, \phi_3, \cdots, \phi_l)$. If such functions exist for some $\Theta = \Theta' \in \omega$, they exist for all $\Theta \in \omega$. Then the region w_0 in W defined by

$$\phi_1(E, \theta_1^0, \vartheta) < k_1(\Phi, \theta_1^0, \vartheta) \quad \textit{and} \quad \phi_1(E, \theta_1^0, \vartheta) > k_2(\Phi, \theta_1^0, \vartheta) \tag{6}$$

is independent of ϑ, and is a region of type B for testing the hypothesis H_0.

Since throughout the proof $\Theta = (\theta_1^0, \vartheta)$, we shall write Θ in place of these symbols to simplify the printing. It is to be understood that *every statement in the proof involving the symbol Θ is asserted for all Θ in ω.*

We suppose first that a type B region w_0 exists in W_+. Then from (a), (c) of definition 1 and assumptions 1^0(a) and 2^0,

$$\int_{w_0} p(E \mid \Theta)\, dW = \alpha, \tag{7}$$

$$\int_{w_0} \phi_1 p(E \mid \Theta)\, dW = 0. \tag{8}$$

Since the value of the integral (7) is independent of ϑ, all its derivatives with respect to $\theta_2, \theta_3, \cdots, \theta_l$ must vanish. This leads [3, pp. 50, 51. Insert k_t before ϕ_t^{-1} in (15)] to

---

[3] By this term we include "product moments."

[4] When I communicated this theorem to Professor Neyman, he informed me it was among the results of a thesis by R. Satô, *Contributions to the theory of testing statistical composite hypotheses*, University of London, 1937, and he kindly sent me a copy of the MS. I decided nevertheless to publish my version of theorem and proof, since for the reasons indicated in section 1 this theory should be available in the literature.

$$(9) \qquad \alpha^{-1} \int_{w_0} \prod_{i=2}^{l} \phi_i^{k_i} p(E \mid \Theta)\, dW = M(k_2, k_3, \cdots, k_l \mid \Theta), \qquad k_i = 0, 1, 2, \cdots,$$

where M is independent of w_0, and thus has the value obtained from (9) by putting $w_0 = W$ and $\alpha = 1$. In particular,

$$(10) \qquad \alpha^{-1} \int_{w_0} \phi_i p(E \mid \Theta)\, dW = 0, \qquad i = 2, 3, \cdots, l.$$

The necessary condition (9) for (7) is also sufficient. Denoting by $\mathcal{E}(f \mid w, \Theta)$ the expected value of a function $f(E, \Theta)$ calculated under the assumption that $E \epsilon w$, equation (9) may be written

$$(11) \qquad \mathcal{E}\left(\prod_{i=2}^{l} \phi_i^{k_i} \mid w_0, \Theta\right) = \mathcal{E}\left(\prod_{i=2}^{l} \phi_i^{k_i} \mid W, \Theta\right).$$

From assumption 5^0(a) it then follows that

$$(12) \qquad Q_0(\Phi \mid w_0, \Theta) = Q_0(\Phi \mid W, \Theta)$$

for almost all Φ. Conversely, (12) implies (11).

In a similar manner we get from (8) with the aid of (9),

$$(13) \qquad \mathcal{E}\left(\phi_1 \prod_{i=2}^{l} \phi_i^{k_i} \mid w_0, \Theta\right) = \mathcal{E}\left(\phi_1 \prod_{i=2}^{l} \phi_i^{k_i} \mid W, \Theta\right).$$

We calculate the moments of the function $Q_1(\Phi \mid w, \Theta)$ to be

$$\mathcal{E}\left(\phi_1 \prod_{i=2}^{l} \phi_i^{k_i} \mid w, \Theta\right),$$

and hence because of 5^0(b), (13) implies

$$(14) \qquad Q_1(\Phi \mid w_0, \Theta) = Q_1(\Phi \mid W, \Theta)$$

almost everywhere in the Φ-space. The pair of conditions (12), (14) are equivalent to the pair (7), (8).

In order that w_0 be a type B region, it is necessary and sufficient that it satisfy (12) and (14) and that

$$P''(w_0 \mid \Theta) \geqq P''(w_1 \mid \Theta)$$

for all w_1 satisfying (12) and (14). The inequality may be transformed with the help of 1^0(a), 2^0, (3), (7), (8), and (10) to

$$\int_{w_0} \phi_1^2 p(E \mid \Theta)\, dW \geqq \int_{w_1} \phi_1^2 p(E \mid \Theta)\, dW,$$

which is equivalent to

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \phi_1^2 p(\phi_1, \Phi \mid w_0, \Theta)\, d\phi_1\, d\phi_2 \cdots d\phi_l$$

$$\geqq \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \phi_1^2 p(\phi_1, \Phi \mid w_1, \Theta)\, d\phi_1\, d\phi_2 \cdots d\phi_l.$$

Sufficient for this is

$$Q_2(\Phi \mid w_0, \Theta) \geqq Q_2(\Phi \mid w_1, \Theta). \tag{15}$$

We note the functions in (12), (14), and (15) are all of the form (4) with $s = 0, 1, 2$, and propose to transform these to integrals over certain portions of the sample space W. First, we write (4) in the form

$$Q_0(\Phi \mid w, \Theta) \int_{-\infty}^{+\infty} \phi_1^s \, p(\phi_1 \mid \Phi, w, \Theta) \, d\phi_1 = Q_0(\Phi \mid w, \Theta) \mathcal{E}(\phi_1^s \mid \Phi, w, \Theta). \tag{16}$$

Next, we consider "surfaces" $S(\Phi, \Theta)$ in W_+, constructed as follows: For any fixed Θ let $D(\Theta)$ be the $l - 1$ dimensional domain of values of $\phi_i(E, \Theta)$, $i = 2, 3, \cdots, l$, for $E \in W_+$. A "surface" $S(\Phi, \Theta)$ is the locus of points E for which

$$\phi_i(E, \Theta) = \phi_i', \text{ a constant}, \qquad i = 2, 3, \cdots, l, \tag{17}$$

the set of constants being in $D(\Theta)$. Over every "surface" we now define a density ρ: Without loss of generality, and to simplify the notation, we shall assume that the non-singular minor postulated in 4^0 contains the minor $(\partial\phi_i/\partial x_j)$, $i = 2, 3, \cdots, l$; $j = 1, 2, \cdots, l - 1$, and denote by $J(E, \Theta)$ its determinant. For E on $S(\Phi, \Theta)$ we define the density

$$\rho(E \mid \Theta) = p(E \mid \Theta) / \mid J(E, \Theta) \mid, \tag{18}$$

and consider "surface" integrals

$$\int_{wS(\Phi,\Theta)} F_s(E, \Theta) \, dx_l \, dx_{l+1} \cdots dx_n, \tag{19}$$

where

$$F_s(E, \Theta) = \phi_1^s(E, \Theta) \rho(E \mid \Theta). \tag{20}$$

A "surface" integral (19) is to be distinguished from an ordinary multiple integral, in that the integrand is not merely a function of $x_l, x_{l+1}, \cdots, x_n$; there may be several points E on the surface with the same values for these coordinates, but different values for the integrand. The integral is to be thought of as follows: The part $wS(\Phi, \Theta)$ of the "surface" $S(\Phi, \Theta)$ is partitioned into pieces ΔS, on each a point E is chosen, and the value of the integrand at E is multiplied by the "area" of the projection (taken non-negative) of ΔS on the $x_l, x_{l+1}, \cdots, x_n$-space. The "surface" integral is the limit of the sum of such products as the norm of the partition approaches zero.

Denoting the integral (19) by $I(s)$ for the moment, we may calculate that for $\Phi \in D(\Theta)$

$$I(s) = I(0)\mathcal{E}(\phi_1^s \mid \Phi, w, \Theta), \qquad I(0) = Q_0(\Phi \mid w, \Theta) P(w \mid \Theta),$$

and hence we see that the right member of (16) is equal to the integral (19) divided by $P(w \mid \Theta)$. The desired relationship between the ordinary integrals (4) and the "surface" integrals (19) is thus

$$(21)\qquad Q_s(\Phi \mid w, \Theta) = \int_{wS(\Phi,\Theta)} F_s(E, \Theta) \prod_{j=l}^{n} dx_j / P(w \mid \Theta).$$

The conditions (12), (14), (15) may now be written

$$(22)\qquad \int_{w_0 S(\Phi,\Theta)} F_s(E, \Theta) \prod_{j=l}^{n} dx_j = \alpha \int_{S(\Phi,\Theta)} F_s(E, \Theta) \prod_{j=l}^{n} dx_j, \qquad s = 0, 1,$$

$$(23)\qquad \int_{w_0 S(\Phi,\Theta)} F_2(E, \Theta) \prod_{j=l}^{n} dx_j \geqq \int_{w_1 S(\Phi,\Theta)} F_2(E, \Theta) \prod_{j=l}^{n} dx_j,$$

if Φ is in the domain $D(\Theta)$, else they are satisfied trivially. w_0 will be a type B region if equations (22) are satisfied for almost all $\Phi \epsilon D(\Theta)$, and if (23) is valid for all w_1 satisfying (22).

We now hold Θ fixed in ω and Φ fixed in $D(\Theta)$, so that $S(\Phi, \Theta)$ is fixed, and the right members of equations (22) have constant values. The proof [5, p. 11] of the lemma of Neyman and Pearson giving sufficient conditions that a region maximize an integral, subject to integral side-conditions, is easily seen to be valid for our "surface" integrals, and a sufficient condition that $w_0S(\Phi, \Theta)$ have the desired property is then that it be defined by

$$(24)\qquad \phi_1^2(E, \Theta) > a_0 + a_1\phi_1(E, \Theta),$$

where a_0, a_1 are independent of E on $S(\Phi, \Theta)$, and are such that equations (22) are satisfied. Since Θ and Φ are fixed, we may permit a_i to be of the nature $a_i = a_i(\Phi, \Theta)$, $i = 1, 2$. Introducing functions $k_1 < k_2$, $k_i = k_i(\Phi, \Theta)$, and defining a_0, a_1 from

$$a_0 = -k_1k_2, \qquad a_1 = k_1 + k_2,$$

the inequality (24) is satisfied if (6) is. Still holding Θ fixed, suppose that k_1, k_2 can be determined for all Φ (hence almost all Φ) in $D(\Theta)$ so that for the part $w_0S(\Phi, \Theta)$ of $S(\Phi, \Theta)$, defined by (6), the equations (22) are satisfied. The parts $w_0S(\Phi, \Theta)$ of "surfaces" then sweep out a "solid" $w_0(\Theta)$ in W_+, defined by (6). If we can similarly determine k_1 and k_2, and hence $w_0(\Theta)$, for every Θ in ω, and if furthermore $w_0(\Theta)$ is independent of Θ, then it is the type B region we seek.

The equations (22) have now served their main purpose, and we return to their equivalents, (12) and (14). For $w_0(\Theta)$ defined by (6)

$$p(\phi_1, \Phi \mid w_0, \Theta) = p(\phi_1, \Phi \mid W, \Theta)/\alpha \quad \text{if} \quad \phi_1 < k_1 \quad \text{or} \quad \phi_1 > k_2,$$

and vanishes otherwise, and hence equations (12) and (14) are equivalent to (5).

The remainder of the proof consists of deducing that k_1, k_2 exist, and that the associated region $w_0(\Theta)$ is independent of Θ, for all $\Theta \epsilon \omega$, from the hypothesis of our theorem that k_1, k_2 exist for some $\Theta = \Theta'$. By 1^0(c), Θ' lies on a line segment L entirely in ω, on which all but one of the nuisance parameters, say θ_2, are constant. Let us vary Θ over L. Then θ_3, θ_4, $\cdots$, θ_l remain fixed and θ_2 varies over an interval I. The equations (2) for $j = 2$ now become

ordinary differential equations in which the independent variable is θ_2, the dependent variables are $\phi_2, \phi_2, \cdots, \phi_l$, and $\theta_1^0, \theta_3, \cdots, \theta_l$ are parameters. A well known existence theorem assures us of the existence of particular solutions u_i and a non-singular (for all θ_2 in I) matrix (u_{ij}) of complementary solutions, $i, j = 2, 3, \cdots, l$, such that the general solution is

$$\phi_i = u_i + \sum_{j=2}^{l} u_{ij} c_j .$$

The u_i are determined by initial conditions for the system (2) with $j = 2$, and the u_{ij} by sets of initial conditions for the corresponding complementary system. Clearly, if these initial conditions are all chosen independent of E, then since the coefficients of the differential equations are all independent of E, the solutions u_i and u_{ij} enjoy the same property. On the other hand, the c_j are independent of θ_2. Hence

$$\phi_i(E, \theta_2) = u_i(\theta_2) + \sum_{j=2}^{l} u_{ij}(\theta_2) c_j(E), \qquad i = 2, 3, \cdots, l. \tag{25}$$

The dependence of the ϕ's, u's and c's on the parameters $\theta_1^0, \theta_3, \cdots, \theta_l$ has not been indicated, since these remain fixed throughout the present calculations.

Let $\mathfrak{D}$ be the $l - 1$ dimensional domain of the values of $c_j(E)$ for $E \epsilon W_+$, and C: $(c_2', c_3', \cdots, c_l')$ be a point in $\mathfrak{D}$, and denote by $S(C)$ the "surface" $c_j(E) = c_j'$. Denote the surface $S(\Phi, \Theta)$ defined in (17) by $S(\Phi, \theta_2)$, and the domain $D(\Theta)$ of Φ by $D(\theta_2)$. Then since $| u_{ij} | \neq 0$, therefore for every $\theta_2 \epsilon I$, every $S(C)$ with $C \epsilon \mathfrak{D}$ is identical with some $S(\Phi, \theta_2)$ with $\Phi \epsilon D(\theta_2)$, and vice versa. From this we conclude for later reference: (A) the functions $c_j(E)$ are constant on every $S(\Phi, \theta_2)$; (B) if θ_2', θ_2'' are any two values in I, then for every $\Phi = \Phi'' \epsilon D(\theta_2'')$ there exists a $\Phi' \epsilon D(\theta_2')$ such that $S(\Phi', \theta_2')$ is identical with $S(\Phi'', \theta_2'')$, and vice versa.

Now let us integrate with respect to θ_2 the equation

$$\partial \log p(E \mid \theta_2)/\partial \theta_2 = \phi_2 = u_2(\theta_2) + \sum_{j=2}^{l} u_{2j}(\theta_2) c_j(E).$$

$$\log p(E \mid \theta_2) = v(\theta_2) + \sum_{j=2}^{l} v_j(\theta_2) c_j(E) + f(E),$$

where $v(\theta_2)$, $v_j(\theta_2)$, $f(E)$, and all new undefined symbols in the sequel have obvious meanings. We get

$$p(E \mid \theta_2) = \bar{v}(\theta_2) \bar{f}(E) \exp\left[\sum_{j=2}^{l} v_j(\theta_2) c_j(E)\right]. \tag{26}$$

Next we differentiate the equations (25) with respect to x_k, and write the result in matrix form,

$$(\partial \phi_i / \partial x_k) = (u_{ij})(\partial c_j / \partial x_k), \qquad i, j = 2, 3, \cdots, l;\ k = 1, 2, \cdots, l - 1.$$

Taking determinants, we have

$$J(E, \theta_2) = J_1(\theta_2) J_2(E). \tag{27}$$

Finally, we shall need to know the nature of the dependence of ϕ_1 on θ_2 and E: From (3),

$$\partial\phi_1/\partial\theta_2 = A_{12}(\theta_2) + B_{121}(\theta_2)\phi_1 + \sum_{j=2}^{l} B_{12k}(\theta_2)\phi_k.$$

Substituting from (25), we get

$$\partial\phi_1/\partial\theta_2 = B_{121}(\theta_2)\phi_1 + A(\theta_2) + \sum_{j=2}^{l} B_j(\theta_2)c_j(E),$$

and integrating,

$$\phi_1(E, \theta_2) = B(\theta_2)\left[\int^{\theta_2} \frac{A(\xi) + \sum_{j=2}^{l} B_j(\xi)c_j(E)}{B(\xi)}\, d\xi + g(E)\right],$$

where

$$B(\theta_2) = \exp\left[\int^{\theta_2} B_{121}(\eta)\, d\eta\right]. \tag{28}$$

Thus

$$\phi_1(E, \theta_2) = \bar{A}(\theta_2) + \sum_{j=2}^{l} \bar{B}_j(\theta_2)c_j(E) + B(\theta_2)g(E). \tag{29}$$

In equations (22) we now use the definitions (20), (18) for the integrands and then substitute (26), (27), (29). As a result we obtain the equality of

$$\int_{w_0S(\Phi,\theta_2)} \frac{\left[\bar{A}(\theta_2) + \sum_{j=2}^{l} \bar{B}_j(\theta_2)c_j(E) + B(\theta_2)g(E)\right]^s \bar{v}(\theta_2)\bar{f}(E) \cdot\exp\left[\sum_{j=l}^{l} v_j(\theta_2)c_j(E)\right]}{|J_1(\theta_2)J_2(E)|} \prod_{j=l}^{n} dx_j$$

and α times the "surface" integral of the same integrand over $S(\Phi, \theta_2)$. Putting first $s = 0$ and then $s = 1$, and employing the previous conclusion (A), we find that the equations (22) are equivalent to

$$\int_{w_0S(\Phi,\theta_2)} \{g^s(E)\bar{f}(E)/|J_2(E)|\} \prod_{j=l}^{n} dx_j = \alpha \int^{\cdot}_{S(\Phi,\theta_2)} \{g^s(E)\bar{f}(E)/|J_2(E)|\} \prod_{j=l}^{n} dx_j, \qquad s = 0, 1. \tag{30}$$

Again using the expression (29) for ϕ_1, and noting from (28) that $B(\theta_2) > 0$, we may write the inequality (6) in the form

$$g(E) < \kappa_1(\Phi, \theta_2) \quad \text{and} \quad g(E) > \kappa_2(\Phi, \theta_2), \tag{31}$$

where

$$\kappa_i(\Phi, \theta_2) = \left[k_i(\Phi, \theta_1^0, \vartheta) - \bar{A}(\theta_2) - \sum_{j=2}^{l} \bar{B}_j(\theta_2)c_j(E)\right] \Big/ B(\theta_2). \tag{32}$$

It follows from our hypothesis that for $\theta_2 = \theta_2'$ (the θ_2 coordinate of Θ') and any $\Phi \epsilon D(\theta_2')$, functions $\kappa_i(\Phi, \theta_2')$ exist such that for the part $w_0S(\Phi, \theta_2')$ of $S(\Phi, \theta_2')$, defined by (31), equations (30) are satisfied. The region $w_0(\Theta')$ is "swept out" by $w_0S(\Phi, \theta_2')$ as Φ ranges over $D(\theta_2')$. Now let Θ'' be any other $\Theta \epsilon L$, call its θ_2 coordinate θ_2'', let Φ'' be any $\Phi \epsilon D(\theta_2'')$, and consider the possibility of finding $\kappa_i(\Phi'', \theta_2'')$ such that on the part $w_0S(\Phi'', \theta_2'')$ of $S(\Phi'', \theta_2'')$, defined by (31), equations (30) are satisfied. From the conclusion (B), $S(\Phi'', \theta_2'')$ is identical with $S(\Phi', \theta_2')$ for a suitably chosen $\Phi' \epsilon D(\theta_2')$. Hence if we take $\kappa_i(\Phi'', \theta_2'') = \kappa_i(\Phi', \theta_2')$, then $w_0S(\Phi'', \theta_2'')$ becomes identical with $w_0S(\Phi', \theta_2')$ where equations (30) are already satisfied. Letting Φ'' range over $D(\theta_2'')$, every $w_0S(\Phi'', \theta_2'')$ thus determined becomes identical with some $w_0S(\Phi', \theta_2')$, and vice versa, by (B). Thus the region $w_0(\Theta'')$ "swept out" is identical with $w_0(\Theta')$. This process defines $\kappa_i(\Phi, \theta_2)$ for all $\theta_2 \epsilon I$ and $\Phi \epsilon D(\theta_2)$, and hence determines $k_i(\Phi, \theta_1^0, \vartheta)$ from (32). We now have functions $k_i(\Phi, \theta_1^0, \vartheta)$, $k_1 < k_2$, satisfying (5), and corresponding regions $w_0(\Theta)$ independent of Θ, for all $\Theta \epsilon L$. To conclude the proof, we use $1^0(c)$ to reach any point Θ in ω from Θ' by a path consisting of a finite number of segments like L on which only one of the nuisance parameters varies. The definitions of $k_i(\Phi, \theta_1^0, \vartheta)$ are continued along this path as above and the region $w_0(\Theta)$ is seen to be independent of Θ for all Θ in ω.

The following theorem may be regarded as a generalization of one by Neyman [6, p. 33] giving sufficient conditions that a type A region be also of type A_1:

THEOREM 2. *Suppose the assumption* $1^0(b)$ *holds for all* $\Theta \epsilon \Omega$. *Denote* $\phi_i(E, \theta_1^0, \vartheta)$ *by* ϕ_i^0 *and let* $R(\vartheta)$ *be the domain of values of* $\phi_1^0, \phi_2^0, \cdots, \phi_l^0$ *for* $E \epsilon W_+$ *and* $\Theta \epsilon \omega$. *Then a sufficient condition that a region* w_0 *of type* B, *found by application of theorem* 1, *be also of type* B_1 *is that for all* $\Theta \epsilon \Omega$ *and all* $E \epsilon W_+$

$$p(E \mid \theta_1, \vartheta) = p(E \mid \theta_1^0, \vartheta) g(\phi_1^0, \phi_2^0, \cdots, \phi_l^0; \theta_1^0; \theta_1, \vartheta), \tag{33}$$

where $g(y_1, y_2, \cdots, y_l; \theta_1^0; \theta_1, \vartheta)$ *is a function such that* $\partial^2 g/\partial y_1^2 > 0$ *for all* $y_1, y_2, \cdots, y_l$ *in* $R(\vartheta)$ *and* $\Theta \epsilon \Omega - \omega$.

For the w_0 satisfying the sufficient conditions of theorem 1, the conditions (a), (b′), (c) of definition 2 are satisfied, and it remains only to verify the condition (d′). The regions w_1 admitted for comparison in (d), as well as w_0, must satisfy the equations (22) since these are equivalent to the conditions (a), (c). We recall that $\Theta = (\theta_1^0, \vartheta)$ in equations (22) and rewrite them in a notation better adapted to our present considerations:

$$\begin{aligned} &\int_{w_0S(\Phi^0,\theta_1^0,\vartheta)} [\phi_1^0]^s \{p(E \mid \theta_1^0, \vartheta)/\lvert J(E, \theta_1^0, \vartheta)\rvert\} \prod_{j=l}^{n} dx_j \\ &\qquad = \alpha \int_{S(\Phi^0,\theta_1^0,\vartheta)} [\phi_1^0]^s \{p(E \mid \theta_1^0, \vartheta)/\lvert J(E, \theta_1^0, \vartheta)\rvert\} \prod_{j=l}^{n} dx_j, \qquad s = 0, 1 \end{aligned} \tag{34}$$

where $\Phi^0 = (\phi_2^0, \phi_3^0, \cdots, \phi_l^0) \epsilon D(\theta_1^0, \vartheta)$.

To express the condition (d) in a convenient way, we now "shred" the regions w_0, w_1 of (d) for every θ_1 by means of the same "surfaces" we have been using

for $\theta_1 = \theta_1^0$: For any w in W_+, $\Theta \epsilon \Omega$, and $\Phi^0 \epsilon D(\theta_1^0, \vartheta)$ we define a "surface" integral

$$I(\Phi^0, w \mid \theta_1, \vartheta) = \int_{wS(\Phi^0, \theta_1^0, \vartheta)} \{p(E \mid \theta_1, \vartheta)/\mid J(E, \theta_1^0, \vartheta) \mid \} \prod_{j=l}^{n} dx_j .$$

Then

$$P(w \mid \theta_1, \vartheta) = \int \cdots \int_{D(\theta_1^0, \vartheta)} I(\Phi^0, w \mid \theta_1, \vartheta) \, d\phi_2^0 \, d\phi_3^0 \cdots d\phi_l^0,$$

and a sufficient condition for (d) is

$$I(\Phi^0, w_0 \mid \theta_1, \vartheta) \geqq I(\Phi^0, w_1 \mid \theta_1, \vartheta) \tag{35}$$

for all $\Theta \epsilon \Omega$ and all $\Phi^0 \epsilon D(\theta_1^0, \vartheta)$.

Again applying the lemma of Neyman and Pearson to the integrands of the "surface" integrals in (34) and (35), we find that a sufficient condition that our region w_0 be of type B_1 is that there exist functions $b_i(\Phi^0, \theta_1^0, \theta_1, \vartheta)$, $i = 1, 2$, such that

$$p(E \mid \theta_1, \vartheta) > p(E \mid \theta_1^0, \vartheta)[b_0 + b_1\phi_1^0(E, \theta_1^0, \vartheta)]$$

if and only if $E \epsilon w_0$. Employing (33), we may replace this inequality by

$$g(\phi_1^0, \Phi^0; \theta_1^0; \theta_1, \vartheta) > b_0 + b_1\phi_1^0 . \tag{36}$$

Define b_0, b_1 from

$$g(k_i, \Phi^0; \theta_1^0; \theta_1, \vartheta) = b_0 + b_1 k_i , \qquad i = 1, 2,$$

where $k_i = k_i(\Phi^0, \theta_1^0, \vartheta)$. Since $k_1 < k_2$, these equations have unique solutions b_0, b_1. Now hold Φ^0, θ_1, ϑ all fixed ($\theta_1 \neq \theta_1^0$) and consider the graphs of the members of (36) as functions of ϕ_1^0. From our definition of b_0, b_1, these graphs intersect at k_1, k_2. But by hypothesis, the graph of the left member is everywhere concave up, and hence for $k_1 < \phi_1^0 < k_2$, it lies below the linear graph of the right member, and for $\phi_1^0 < k_1$ and $\phi_1^0 > k_2$, it lies above. That is (36) is true if and only if $E \epsilon w_0$.

5. Appendix on the moment problem. Easily applied criteria [8] are available for the moment problem of assumption $5^0(a)$. The moment problem $5^0(b)$ is much more difficult, however, because the function to be determined by its moments is not of constant sign. Below we offer a proof that the solutions of both problems $5^0(a)$ and $5^0(b)$ are unique in the important case where $p(E \mid \Theta)$ is a multivariate normal p.d.f. and $\phi_1, \phi_2, \cdots, \phi_l$ are polynomials in x_1, $x_2, \cdots, x_n$ of degree $\leqq 2$ and not necessarily homogeneous. Since Θ is held fixed, we will not indicate dependence on Θ, nor will the dependence of various functions on s be indicated, since $s = 0$ or else 1 throughout.

Let w_1, w_2 be any two regions, $\alpha_j = P(w_j) \neq 0$, for which the moments of $Q_s(\Phi \mid w_1)$ and $Q_s(\Phi \mid w_2)$ are the same. To prove the equality (almost every-

where) of these two functions it suffices to prove that their Fourier transforms are identical [7, theorem 61]. Suppressing the customary multiple of $\sqrt{2\pi}$, the Fourier transform of $Q_s(\Phi \mid w_j)$ is

$$\Psi_j(\mathbf{t}) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} e^{i\mathbf{t}\cdot\Phi} Q_s(\Phi \mid w_j)\, d\phi_2 \cdots d\phi_l,$$

where $\mathbf{t}$ is the vector $(t_2, t_3, \cdots, t_l)$ and $\mathbf{t}\cdot\Phi = t_2\phi_2 + \cdots + t_l\phi_l$. From (4) we get

$$\begin{aligned}\Psi_j(\mathbf{t}) &= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} e^{i\mathbf{t}\cdot\Phi} \phi_1^s\, p(\phi_1, \Phi \mid w_j)\, d\phi_1\, d\phi_2 \cdots d\phi_l \\ &= \mathcal{E}(e^{i\mathbf{t}\cdot\Phi} \phi_1^s \mid w_j) \\ &= \frac{1}{\alpha_j} \int_{w_j} e^{i\mathbf{t}\cdot\Phi} \phi_1^s(E) p(E)\, dW.\end{aligned}$$

A device of Cramér and Wold [8] for reducing the dimensionality of the problem now suggests itself. Let z be a scalar variable and consider $\psi_j(z \mid \mathbf{t}) = \Psi_j(z\mathbf{t})$ for fixed $\mathbf{t}$ as a function of z. Obviously if for every fixed $\mathbf{t}$, $\psi_1(z \mid \mathbf{t}) = \psi_2(z \mid \mathbf{t})$, then $\Psi_1(\mathbf{t}) = \Psi_2(\mathbf{t})$, and we are through. We propose to prove the former equality by showing first that ψ_j is an analytic function of z for all real z and secondly that the coefficients of the power series for ψ_1 and ψ_2 in powers of z are equal. Holding $\mathbf{t}$ fixed now, $\xi = \mathbf{t}\cdot\Phi$ is a polynomial of degree $\leqq 2$, and

$$\psi_j(z \mid \mathbf{t}) = \frac{1}{\alpha_j} \int_{w_j} e^{iz\xi} \phi_1^s\, p\, dW. \tag{37}$$

By our assumption of normality,

$$p = C \exp\left[-\sum_{\kappa,\nu=1}^{n} a_{\kappa\nu} y_\kappa y_\nu \right], \qquad y_\kappa = x_\kappa - \mu_\kappa,$$

where the matrix $(a_{\kappa\nu})$ is positive definite. To prove the analyticity of ψ_j for any real $z = z_0$, let $z = z_0 + \zeta$, and restrict ζ to real values. Substitute in (37)

$$e^{i\zeta\xi} = \sum_{q=0}^{m-1} \frac{(i\zeta\xi)^q}{q!} + \frac{(i\zeta\xi)^m}{m!} f_m(\zeta\xi),$$

where $|f_m(\zeta\xi)| \leqq 1$. Then

$$\psi_j(z_0 + \zeta \mid \mathbf{t}) = \sum_{q=0}^{m-1} \frac{(i\zeta)^q}{q!\,\alpha_j} \int_{w_j} e^{iz_0\xi} \xi^q \phi_1^s\, p\, dW + R_{jm}(z_0, \zeta),$$

where

$$R_{jm} = \frac{(i\zeta)^m}{m!\,\alpha_j} \int_{w_j} e^{iz_0\xi} f_m(\zeta\xi) \xi^m \phi_1^s\, p\, dW,$$

and all integrands are absolutely integrable over W. Let σ be the sphere of unit radius with center at $(\mu_1, \mu_2, \cdots, \mu_n)$ in W and write

$$R_{jm} = \frac{(i\zeta)^m}{m!\,\alpha_j}\left[\int_{w_j\sigma} + \int_{w_j - w_j\sigma}\right].$$

Call the two terms of the right member R'_{jm} and R''_{jm},

$$R_{jm} = R'_{jm} + R''_{jm}.$$

$$|R'_{jm}| \leqq \frac{|\zeta|^m}{m!\,\alpha_j}\int_{w_j\sigma} |\xi^m \phi_1^s|\, p\, dW.$$

Let $M = \max|\xi|$, $M_1 = \max|\phi_1^s|$, for $E \in \sigma$. Then

$$|R'_{jm}| \leqq \frac{M_1 |M\zeta|^m}{m!\,\alpha_j}\int_\sigma p\, dW \leqq M_1|M\zeta|^m/m!\,\alpha_j.$$

Hence $R'_{jm} \to 0$ for all real ζ as $m \to \infty$.

$$|R''_{jm}| \leqq \frac{|\zeta|^m}{m!\,\alpha_j}\int_{W-\sigma} |\xi^m \phi_1^s|\, p\, dW.$$

Let $r = \left(\sum_{\kappa=1}^{n} y_\kappa^2\right)^{\frac{1}{2}}$, and M_2, M_3 be the sums of the absolute values of the coefficients of the polynomials ϕ_1^s, ξ, respectively, when expanded in powers of y_κ. Then for $E \in W - \sigma$, $|\phi_1^s| \leqq M_2 r^2$, $|\xi| \leqq M_3 r^2$, $p \leqq C \exp(-\lambda r^2)$, where $\lambda > 0$ is the smallest characteristic root of $(a_{\kappa\nu})$. Hence

$$|R''_{jm}| \leqq \frac{CM_2|M_3\zeta|^m}{m!\,\alpha_j}\int_{W-\sigma} r^{2m+2} e^{-\lambda r^2}\, dW.$$

Integrating over spherical shells concentric with σ, $dW = M_4 r^{n-1}\, dr$, and

$$|R''_{jm}| \leqq \frac{CM_2 M_4|M_3\zeta|^m}{m!\,\alpha_j}\int_1^\infty r^{2m+n+1} e^{-\lambda r^2}\, dr \leqq \frac{CM_2 M_4|M_3\zeta|^m}{m!\,\alpha_j}\int_0^\infty.$$

If we evaluate the last integral in terms of a Gamma function and employ Stirling's formula we easily find that for $M_3|\zeta| < \lambda$, $R''_{jm} \to 0$. The convergence of R_{jm} to zero for *real* ζ, $|\zeta| < \lambda/M_3$, is sufficient to insure the analyticity of ψ_j.

Now let $z_0 = 0$. Then the coefficient of ζ^q in the power series for ψ_j is

$$\frac{i^q}{q!\,\alpha_j}\int_{w_j} (t_2\phi_2 + \cdots + t_l\phi_l)^q \phi_1^s\, p\, dW,$$

a linear combination (the same for $j = 1, 2$) of the q-th order moments of $Q_s(\Phi \mid w_j)$, and hence corresponding coefficients for ψ_1 and ψ_2 are equal.

REFERENCES

[1] J. Neyman, "Sur la vérification des hypothèses statistiques composées," *Bull. Soc. Math. France*, Vol. 63 (1935), pp. 246–266.

[2] J. Neyman, "Outline of a theory of statistical estimation based on the classical theory of probability," *Phil. Trans. Roy. Soc. London*, A, Vol. 236 (1937), pp. 333–380.

[3] J. Neyman, "On a statistical problem arising in routine analyses and in sampling inspections of mass production," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 46–76.

[4] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Phil. Trans. Roy. Soc. London*, A, Vol. 231 (1933), pp. 289–337.

[5] J. Neyman and E. S. Pearson, "Contributions to the theory of testing statistical hypotheses: Part I," *Stat. Res. Mem.*, Vol. 1 (1936), pp. 1–37.

[6] J. Neyman and E. S. Pearson, "Contributions to the theory of testing statistical hypotheses: Part II," *Stat. Res. Mem.*, Vol. 2 (1938), pp. 25–36.

[7] S. Bochner, *Vorlesungen über Fouriersche Integrale*, Leipzig, 1932.

[8] H. Cramér and H. Wold, "Some theorems on distribution functions," *Jour. London Math. Soc.*, Vol. 11 (1936), pp. 290–294.

ON THE PROBLEM OF MULTIPLE MATCHING

By I. L. Battin

Drew University

1. Introduction. The problem of determining the distribution of the number of "hits" or "matchings" under random matching of two decks of cards has received attention from a number of authors within the last few years. In 1934 Chapman [2] considered pairings between two series of t elements each, and later [3] generalized the problem to series of u and $t(\leq u)$ elements respectively. In the same paper he also considered the distribution of the mean number of correct matchings resulting from n independent trials, and gave a method, and tables, for determining the significance of any obtained mean. In 1937 Bartlett [1] considered matchings of two decks of cards, using a number of interesting moment generating functions. In 1937 Huntington [12, 13] gave tables of probabilities for matchings between decks with the compositions (5^5), (4^4), and (3^3), where (s^t) denotes a deck consisting of s of each of t kinds of cards. More generally $(s_1s_2 \cdots s_t)$ denotes s_1 cards of the first kind, s_2 of the second, etc. Sterne [16] has given the first four moments of the frequency distribution for the (5^5) case and has fitted a Pearson Type I distribution function to the distribution. Sterne obtained his results by considering the probabilities in a 5×5 contingency table. He also considered the 4×4 and 3×3 cases. In 1938 Greville [7] gave a table of the exact probabilities for matchings between two decks of compositions (5^5). Greenwood [4] obtained the variance of the distribution of hits for matchings between two decks having the respective compositions (s^t) and $(s_1s_2 \cdots s_t)$ with $s_1 + s_2 + \cdots + s_t = st = n$, and where it is not necessary that all the s's should be different from zero. Earlier Wilks [19] had considered the same problem for $t = 5$ and $n = 25$.

In a very interesting paper Olds [15] in 1938 used permanents to express a moment generating function suitable for the problem in question. He obtained factorial moments and the first four ordinary moments about the mean, first for two decks with composition (4^2), and then for two decks of composition (s^t). In 1938 Stevens [17] considered a contingency table in connection with matchings between two sets of n objects each, and gave the means, variances, and covariances of the single cell entries and various sub-totals of the cell entries. Stevens [18] also gave a treatment of the problem of matchings between two decks which was based on elementary considerations. In 1940 Greenwood [6] gave the first four moments of the distribution of hits between two decks of any composition whatever, generalizing the problem which had been treated earlier by Olds [15]. Finally in 1941, Greville [8] gave the exact distribution of hits for matchings between two decks of arbitrary composition. He also considered the problem from the standpoint of a contingency table, as had been done earlier by Stevens.

In 1939 Kullback [14] considered matchings between two sequences obtained by drawing at random a single element in turn from each of n urns U_i containing elements of r types E_j in the respective proportions p_{ij}. He showed that if the process of drawing were indefinitely repeated the distribution of hits would be that of a Poisson series.

The work which has been done thus far applies to the problem of matching two decks of cards. In the present paper a method is developed for obtaining the moments of the distribution of hits for matchings between three or more decks of cards of arbitrary composition.

2. Matchings between two Decks of cards. In the present paper it will be convenient to take as the point of departure the method used by Wilks [19] in his treatment of the problem of hits occurring under random matching of two decks of 25 cards each, namely a target deck with composition (5^5) and a matching deck with composition (s_i), $i = 1, 2, 3, \cdots, 5$, $\sum_i s_i = 25$. He showed that

$$(1)\quad \phi = \frac{1}{\begin{bmatrix} 25 \\ s_i \end{bmatrix}} (x_1 e^\theta + x_2 + \cdots + x_5)^5 (x_1 + x_2 e^\theta + x_3 + \cdots + x_5)^5 \cdots (x_1 + x_2 + \cdots + x_5 e^\theta)^5$$

where,

$$\begin{bmatrix} 25 \\ s_i \end{bmatrix} \equiv \frac{25!}{s_1!\, s_2! \cdots s_5!},$$

is a suitable generating function for obtaining the moments of the distribution. In fact, if we define an operator $K_{s_1 s_2 \cdots s_5}$ as

$$(2)\qquad K_{s_1 s_2 \cdots s_5} u \;\equiv\; \text{coefficient of } x_1^{s_1} x_2^{s_2} \cdots x_5^{s_5} \text{ in } u,$$

where $u = u(x_1, x_2, \cdots, x_5)$, and if h denotes the number of hits, then for $r = 1, 2, \cdots, 5$,

$$(3)\qquad P(h = r) = \text{coefficient of } e^{r\theta} \text{ in } K_{s_1 s_2 \cdots s_5} \phi$$

And it is readily seen that

$$(4)\qquad E(h^p) = K_{s_1 s_2 \cdots s_5} \left. \frac{\partial^p \phi}{\partial \theta^p} \right|_{\theta = 0}.$$

Wilks' ϕ function involves a particular order for the target deck. If we are to generalize and obtain moments for matchings between more than two decks, it is obvious that we must devise a procedure which will, in the case of two decks, be perfectly symmetrical and not require that one deck be given a preferred status. In the case of two decks this is readily accomplished by the use of Kronecker deltas, and in the case of three or more decks by the use of obvious generalizations of these deltas.

For two decks of 25 cards each with compositions (5^5) we need only let

$$(5) \qquad \phi = (x_i\, y_j\, e^{\delta_{ij}\theta})^{25} \equiv \left(\sum_{i,j=1}^{5} x_i\, y_j\, e^{\delta_{ij}\theta}\right)^{25}$$

where $\delta_{ii} = 1$; $\delta_{ij} = 0$, $i \neq j$.
Then, if

(6) $K_{n_{11}n_{12}\cdots n_{15}\cdot n_{21}n_{22}\cdots n_{25}}\, u \equiv$ coefficient of $x_1^{n_{11}}x_2^{n_{12}} \cdots x_5^{n_{15}}y_1^{n_{21}}y_2^{n_{22}} \cdots y_5^{n_{25}}$ in n, where $u = u(x_1, x_2, \cdots, x_5, y_1, y_2, \cdots, y_5)$, it readily follows that

$$(7) \qquad E(h^p) = \frac{K_{55555\cdot 55555} \left.\dfrac{\partial^p \phi}{\partial \theta^p}\right|_{\theta=0}}{K_{55555\cdot 55555}\, \phi \,|_{\theta=0}}.$$

More generally, for two decks of n cards each, the cards being of k types, and the decks having compositions $(n_{11}, n_{12}, \cdots, n_{1k})$, $(n_{21}, n_{22}, \cdots, n_{2k})$ respectively, we let

$$(8) \qquad \phi = u^n \equiv (x_i\, y_j\, e^{\delta_{ij}\theta})^n \equiv \left(\sum_{i,j=1}^{k} x_i\, y_j\, e^{\delta_{ij}\theta}\right)^n.$$

The factors of ϕ are in one-to-one correspondence with the n events of dealing a card from each of the two decks. The values which can be assumed by the subscripts i and j are in one-to-one correspondence with the k types of cards. The symbol x_i corresponds to the first deck, y_j to the second, the subscripts i and j corresponding to the different types of cards in each deck. The expansion of ϕ consists of all products which can be formed by choosing one and only one pair $x_\alpha y_\beta$ from each factor of ϕ as a factor of the product. In forming any term of ϕ, choosing $x_\alpha y_\alpha$ from any factor of ϕ corresponds to dealing a card of type α from both decks, and introduces e^θ into the coefficient of the term. Choosing $x_\alpha y_\beta$ from any factor corresponds to dealing a card of type α from the first deck, β from the second. If $\alpha \neq \beta$, then, since $\delta_{ij} = 0$, $i \neq j$, e^θ is not introduced into the coefficient. Therefore in the coefficient of any term of ϕ, e^θ will be raised to a power, say s, which is equal to the number of factors of ϕ from which pairs $x_\alpha y_\alpha$ have been chosen.

The total number of ways in which the term

$$x_1^{n_{11}}x_2^{n_{12}} \cdots x_k^{n_{1k}}y_1^{n_{21}}y_2^{n_{22}} \cdots y_k^{n_{2k}}$$

can arise is equal to the number of ways in which two decks of types (n_{1i}), (n_{2j}) respectively can be dealt, (where $(n_{1i}) \equiv (n_{11}n_{12} \cdots n_{1k})$ and similarly for (n_{2j})). But this is given by

$$(9) \qquad \begin{aligned} K_{n_{11}n_{12}\cdots n_{1k}\cdot n_{21}n_{22}\cdots n_{2k}}\, \phi\,|_{\theta=0} &= K_{n_{11}n_{12}\cdots n_{1k}\cdot n_{21}n_{22}\cdots n_{2k}} \left(\sum_{i=1}^{k} x_i\right)^n \left(\sum_{i=1}^{k} y_i\right)^n \\ &= K_{n_{11}n_{12}\cdots n_{1k}} \left(\sum_{i=1}^{k} x_i\right)^n K_{n_{21}n_{22}\cdots n_{2k}} \left(\sum_{i=1}^{k} y_i\right)^n \\ &= \begin{bmatrix} n \\ n_{1i} \end{bmatrix} \begin{bmatrix} n \\ n_{2j} \end{bmatrix}. \end{aligned}$$

The coefficient of $e^{s\theta}$ in $K_{n_{11}n_{12}\cdots n_{1k}\cdot n_{21}n_{22}\cdots n_{2k}}\phi$ is the total number of ways in which the term $x_1^{n_{11}}x_2^{n_{12}}\cdots x_k^{n_{1k}}y_1^{n_{21}}y_2^{n_{22}}\cdots y_k^{n_{2k}}$ can be formed subject to the restriction that pairs x_iy_j with $i = j$ are chosen from s of the factors of ϕ. But this is precisely the number of ways in which the two decks can be dealt so that there will be s hits. Hence if, as above, h is the number of hits, the probability that $h = s$, assuming all permutations in each deck to be equally likely, is given by

$$(10)\qquad P(h = s) = \frac{\text{coefficient of } e^{s\theta} \text{ in } K_{n_{11}n_{12}\cdots n_{1k}\cdot n_{21}n_{22}\cdots n_{2k}}\phi}{K_{n_{11}n_{12}\cdots n_{1k}\cdot n_{21}n_{22}\cdots n_{2k}}\phi\,\big|_{\theta=0}}.$$

Since this is true for all values of s it follows that

$$(11)\qquad E(h^p) = \frac{K_{n_{11}n_{12}\cdots n_{1k}\cdot n_{21}n_{22}\cdots n_{2k}}\dfrac{\partial^p\phi}{\partial\theta^p}\Big|_{\theta=0}}{K_{n_{11}n_{12}\cdots n_{1k}\cdot n_{21}n_{22}\cdots n_{2k}}\phi\,\big|_{\theta=0}}.$$

Since

$$\frac{\partial\phi}{\partial\theta}\Big|_{\theta=0} = nu^{n-1}\left[\sum_{i,j=1}^{k}\delta_{ij}x_iy_je^{\delta_{ij}\theta}\right]\Bigg|_{\theta=0} = n\left[\sum_{i=1}^{k}x_iy_ie^{\theta}\right]u^{n-1}\Bigg|_{\theta=0}$$

$$= n\left[\sum_{i=1}^{k}x_iy_i\right]\left(\sum_{i=1}^{k}x_i\right)^{n-1}\left(\sum_{j=1}^{k}y_j\right)^{n-1}$$

we have at once

$$E(h) = \frac{n}{\begin{bmatrix} n \\ n_{1i}\end{bmatrix}\begin{bmatrix} n \\ n_{2j}\end{bmatrix}}\sum_{i=1}^{k}K_{n_{11}n_{12}\cdots n_{1i-1}(n_{1i}-1)n_{1i+1}\cdots n_{1k}}\left(\sum_{i=1}^{k}x_i\right)^{n-1}$$

$$\cdot K_{n_{21}n_{22}\cdots n_{2i-1}(n_{2i}-1)n_{2i+1}\cdots n_{2k}}\left(\sum_{j=1}^{k}y_j\right)^{n-1}$$

$$(12)\qquad = \frac{n}{\begin{bmatrix} n \\ n_{1i}\end{bmatrix}\begin{bmatrix} n \\ n_{2j}\end{bmatrix}}\sum_{i=1}^{k}\left[\frac{(n-1)!}{n_{11}!\cdots n_{1i-1}!(n_{1i}-1)!\,n_{1i+1}!\cdots n_{1k}!}\right]$$

$$\cdot\left[\frac{(n-1)!}{n_{2i}!\cdots n_{2i-1}!(n_{2i}-1)!\,n_{2i+1}!\cdots n_{2k}!}\right]$$

$$= \sum_{i=1}^{k}\frac{n_{1i}n_{2i}}{n}.$$

It is an equally straightforward matter to show that

$$(13)\qquad E(h^2) = \sum_i\left[\frac{n_{1i}n_{2i}}{n} + \frac{n_{1i}(n_{1i}-1)n_{2i}(n_{2i}-1)}{n(n-1)}\right] + \sum_{i\neq j}\frac{n_{1i}n_{1j}n_{2i}n_{2j}}{n(n-1)}$$

and that

$$(14)\qquad \sigma_h^2 = \sum_i\left[\frac{n_{1i}n_{2i}}{n} - \frac{n_{1i}^2n_{2i}^2}{n^2} + \frac{n_{1i}^{(2)}\,n_{2i}^{(2)}}{n^{(2)}}\right] + \sum_{i\neq j}\frac{n_{1i}n_{1j}n_{2i}n_{2j}}{n^2(n-1)}.$$

Evidently any of the n_{1i} and n_{2j} may be zero, provided only that $\sum_{i=1}^{k} n_{1i} = \sum_{j=1}^{k} n_{2j} = n$. The case of two decks with unequal numbers of cards m and n, $(m < n)$, is readily handled by substituting for the smaller deck one obtained by adding n-m "blank" cards—that is, cards of any type not already appearing in either deck, as indicated by Greville [8], who however obtained his results by considering a preferred order for one of the decks.

EXAMPLE 1. In the case of the decks treated by Wilks [19], $n = 25$, $k = 5$, $n_{1i} = n_{2j} = 5$. Hence from (12)

$$E(h) = \sum_{i=1}^{5} \left\{\frac{5\cdot 5}{25}\right\} = 5,$$

and from (14)

$$\sigma_h^2 = \sum_{i=1}^{5} \left\{\frac{5\cdot 5}{25} - \frac{25\cdot 25}{(25)^2} + \frac{5\cdot 4\cdot 5\cdot 4}{25\cdot 24}\right\} + \sum_{\substack{i,j=1\\ i\neq j}}^{5} \frac{5\cdot 5\cdot 5\cdot 5}{(25)^2 24}$$

$$= \sum_{i=1}^{5} \frac{16}{24} + \sum_{\substack{i,j=1\\ i\neq j}}^{5} \frac{1}{24} = 4\frac{1}{6}.$$

EXAMPLE 2. Suppose we have two decks as shown by the scheme

	Type of card 1	2	3	4	5	Total of all types
No. in deck A	5	7	8	0	0	20
No. in deck B	0	3	4	6	2	15

Here deck B has five fewer cards than deck A. Hence we must presume that there are six types of cards in all, and that the decks have the respective distributions (578000) and (034625). We then have at once

$$E(h) = \sum_{i=1}^{6} \frac{n_{1i}\, n_{2i}}{n} = \frac{1}{20}[0 + 3\cdot 7 + 4\cdot 8 + 0 + 0 + 0]$$

$$= 2.65$$

$$\sigma_h^2 = \sum_{i=1}^{6} \left\{\frac{n_{1i}\, n_{2i}}{n} - \frac{n_{1i}^2\, n_{2i}^2}{n^2} + \frac{n_{1i}^{(2)}\, n_{2i}^{(2)}}{n^{(2)}}\right\} + \sum_{\substack{i,j=1\\ i\neq j}}^{6} \frac{n_{1i}\, n_{2i}\, n_{1j}\, n_{2j}}{n^2(n-1)}$$

$$= 2.65 - \frac{1}{400}\{3^2\cdot 7^2 + 4^2\cdot 8^2\} + \frac{1}{20\cdot 19}\{3\cdot 2\cdot 7\cdot 6 + 4\cdot 3\cdot 8\cdot 7\}$$

$$+ \frac{1}{400\cdot 19}\{3\cdot 7\cdot 4\cdot 8 + 4\cdot 8\cdot 3\cdot 7\}$$

$$= 2.49.$$

3. Matchings between three decks. Let the three decks be of types $(n_{11}n_{12} \cdots n_{1q})$, $(n_{21}n_{22} \cdots n_{2q})$, $(n_{31}n_{32} \cdots n_{3q})$ respectively, with $\sum_{i=1}^{q} n_{1i} = \sum_{j=1}^{q} n_{2j} = \sum_{k=1}^{q} n_{3k} = n$, and consider the function

$$(15) \qquad \phi = \left[\sum_{i,j,k=1}^{q} x_i y_j z_k e^{\delta_{ijk}\theta_{123} + \delta_{ij}\theta_{12} + \delta_{ik}\delta_{13} + \delta_{jk}\theta_{23}} \right]^n \equiv u^n,$$

where

$$(16) \qquad \delta_{iii} = 1, \qquad \delta_{ijk} = 0 \qquad i,\ j,\ k \text{ not all equal},$$

and the other deltas are the usual Kronecker symbols.

Each factor of ϕ corresponds to one deal from each of the three decks. The symbols x, y, and z correspond respectively to cards in the first, second, and third decks. The subscripts $i, j, k, = 1, 2, \cdots, q$ correspond to the types of cards—there being q distinct types.

Choosing $x_\alpha y_\alpha z_\alpha$ from a factor of ϕ corresponds to a deal in which a card of type α is dealt from all three decks, and introduces $e^{\theta_{123}+\theta_{12}+\theta_{13}+\theta_{23}}$ into the coefficient of the corresponding term in the expansion of ϕ. Similarly, choosing $x_\alpha y_\alpha z_\beta$, $\beta \neq \alpha$, corresponds to a hit between the first and second decks, and introduces $e^{\theta_{12}}$ into the coefficient. Similarly choosing $x_\alpha y_\beta z_\alpha$ introduces $e^{\theta_{13}}$; $x_\beta y_\alpha z_\alpha$ introduces $e^{\theta_{23}}$. Choosing $x_\alpha y_\beta z_\gamma$, $\alpha \neq \beta \neq \gamma \neq \alpha$ corresponds to a deal with no hits, and introduces no powers of e into the coefficient, since all the δ's are zero.

Let $K_{n_{1i}\cdot n_{2j}\cdot n_{3k}}$ be defined by

$$(17) \qquad K_{n_{1i}\cdot n_{2j}\cdot n_{3k}} u \equiv \text{coefficient of } x_1^{n_{11}} \cdots x_q^{n_{1q}} y_1^{n_{21}} \cdots y_q^{n_{2q}} z_1^{n_{31}} \cdots z_q^{n_{3q}} \text{ in } u.$$

Then the coefficient of $e^{h_{123}\theta_{123}}$ in $K_{n_{1i}\cdot n_{2j}\cdot n_{3k}}\phi \mid_{\theta_{12}=\theta_{13}=\theta_{23}=0}$ is the number of ways in which the cards can be dealt so as to yield precisely h_{123} triples, or hits between all three decks. Similarly the coefficient of $e^{h_{12}\theta_{12}}$ in $K_{n_{1i}\cdot n_{2j}\cdot n_{3k}}\phi \mid_{\theta_{12}=\theta_{13}=\theta_{23}=0}$ is the number of ways in which the cards can be dealt so as to yield precisely h_{12} hits between the first and second decks, with corresponding results for the first and third (h_{13}) and second and third (h_{23}) decks.

By the same reasoning as before then, we have

$$(18) \qquad E(h_{123}^r) = \frac{K_{n_{1i}\cdot n_{2j}\cdot n_{3k}} \left.\dfrac{\partial^r \phi}{\partial \theta_{123}^r}\right|_{\theta\text{'s}=0}}{K_{n_{1i}\cdot n_{2j}\cdot n_{3k}} \phi \mid_{\theta\text{'s}=0}},$$

$$(19) \qquad E(h_{12}^r) = \frac{K_{n_{1i}\cdot n_{2j}\cdot n_{3k}} \left.\dfrac{\partial^r \phi}{\partial \theta_{12}^r}\right|_{\theta\text{'s}=0}}{K_{n_{1i}\cdot n_{2j}\cdot n_{3k}} \phi \mid_{\theta\text{'s}=0}},$$

with similar results for h_{13} and h_{23}. And it is a straightforward matter to show that

$$(20)\quad E(h_{123}) = n \sum_{i=1}^{q} \left(\prod_{\alpha=1}^{3} \frac{n_{\alpha i}}{n} \right)$$

$$(21)\quad \begin{aligned} E(h_{123}^2) = n \sum_{i=1}^{q} \left(\prod_{\alpha=1}^{3} \frac{n_{\alpha i}}{n} \right) + n(n-1) \sum_{i=1}^{q} \left(\prod_{\alpha=1}^{3} \frac{n_{\alpha i}^{(2)}}{n^{(2)}} \right) \\ + n(n-1) \sum_{i,j=1,(i\neq j)}^{q} \left(\prod_{\alpha=1}^{3} \frac{n_{\alpha i} n_{\alpha j}}{n^{(2)}} \right). \end{aligned}$$

$$(22)\quad E(h_{12}) = \frac{1}{n^2} \sum_{i,k=1}^{q} n_{1i} n_{2j} n_{3k}$$

$$(23)\quad E(h_{13}) = \frac{1}{n^2} \sum_{j,k=1}^{q} n_{1k} n_{2j} n_{3k}$$

$$(24)\quad E(h_{23}) = \frac{1}{n^2} \sum_{i,j=1}^{q} n_{1i} n_{2j} n_{3j}$$

$$(25)\quad \begin{aligned} E(h_{12}^2) = \frac{1}{n^2} \sum_{i,k} n_{1i} n_{2i} n_{3k} + \frac{1}{n^2(n-1)^2} \Bigg[\sum_{i,k} n_{1i}^{(2)} n_{2i}^{(2)} n_{3k}^{(2)} \\ + \sum_{i,\, k\neq r} n_{1i}^{(2)} n_{2i}^{(2)} n_{3k} n_{3r} + \sum_{k,\; i\neq l} n_{1i} n_{1l} n_{2i} n_{2l} n_{3k}^{(2)} \\ + \sum_{i\neq l,\, k\neq r} n_{1i} n_{1l} n_{2i} n_{2l} n_{3k} n_{3r} \Bigg] \end{aligned}$$

with corresponding results for other moments. It is understood each summation index takes values from 1 to q.

As before, if the decks do not all have the same total number of cards it is merely necessary to introduce one or more sets of "blank" cards. Thus we would replace decks with the compositions (57800), (03462), (00335) by hypothetical decks (5780000), (0346250), (0033509) and proceed as before.

EXAMPLE 3. For three decks of 25 cards, consisting of five of each of five kinds we have $n = 25$, $n_{\alpha i} = 5$, $\alpha = 1, 2, 3$, $i = 1, 2, \cdots, 5$. Hence

$$E(h_{123}) = 25 \sum_{i=1}^{5} \prod_{\alpha=1}^{3} \frac{5}{25} = 1$$

$$E(h_{123}^2) = 25 \sum_{i=1}^{5} \left(\frac{5}{25}\right)^3 + 25\cdot 24 \sum_{i=1}^{5} \left(\frac{5\cdot 4}{25\cdot 24}\right)^3 + 25\cdot 24 \sum_{\substack{i,j=1\\ i\neq j}}^{5} \left(\frac{5^2}{25\cdot 24}\right)^3$$

$$= 1\frac{47}{48}$$

$$\sigma_{h_{123}}^2 = \frac{47}{48}.$$

$$E(h_{12}) = \frac{1}{(25)^2} \sum_{i,k=1}^{5} 5^3$$

$$= 5$$

$$E(h_{12}^2) = \frac{1}{(25)^2} \sum_{i,k=1}^{5} 5^3 + \frac{1}{(25)^2(24)^2} \Bigg[\sum_{i,k=1}^{5} 5^3 4^3 + \sum_{\substack{i,k,r=1\\ k \neq r}}^{5} 5^4 4^2 + \sum_{\substack{i,l,k=1\\ i\neq l}}^{5} 5^5 4 + \sum_{\substack{i,l,k,r=1\\ i \neq l\\ k\neq r}}^{5} 5^6 \Bigg]$$

$$= 29\tfrac{1}{6},$$

$$\sigma_{h_{12}}^2 = 4\tfrac{1}{6}.$$

with similar results for $E(h_{13})$, $E(h_{23})$, $\sigma_{h_{13}}^2$, and $\sigma_{h_{23}}^2$.

4. Generalization to any number of decks. If the moments of the distribution of hits—doubles, triples, quadruples, ...—in matching any number of decks is desired, these can be obtained by using an obvious generalization of (15). Thus for four decks we would define $\delta_{iiii} = 1$, $\delta_{ijkl} = 0$, i, j, k, l not all equal, and use

$$\phi = \left[\sum_{i,j,k,l=1}^{q} x_i \dot{y}_j z_k w_l e^{\delta_{ijkl}\theta_{1234} + \delta_{ijk}\theta_{123} + \delta_{ijl}\theta_{124} + \cdots + \delta_{ij}\theta_{12} + \cdots + \delta_{kl}\theta_{34}} \right]^n \tag{26}$$

However, it is evident that as the number of decks is increased the summations involved and the manipulation of the (generalized) K operators rapidly become complicated.

5. Application of our moment-generating technique to two-way contingency tables. The moment-generating technique which we have discussed has wider applications than merely to matching problems. As an example of considerable interest we shall consider the contingency problem. Consider the array

$$\begin{array}{c|c} n_{\alpha\beta} & n_{\alpha\cdot} \\ \hline n_{\cdot\beta} & n \end{array} \qquad \begin{array}{c} \alpha = 1, 2, \cdots, r \\ \beta = 1, 2, \cdots, s \\ \sum_{\alpha,\beta} n_{\alpha\beta} = \sum_{\alpha} n_{\alpha\cdot} = \sum_{\beta} n_{\beta\cdot} = n \end{array} \tag{27}$$

and also the function

$$\phi = \prod_{\alpha=1}^{r} (x_\beta e^{\theta_{\alpha\beta}})^{n_{\alpha\cdot}} \equiv \prod_{\alpha=1}^{r} \left(\sum_{\beta=1}^{s} x_\beta e^{\theta_{\alpha\beta}} \right)^{n_{\alpha\cdot}}. \tag{28}$$

If i and j are particular values of α and β respectively, then to the i-th row of the array corresponds the product $(x_\beta e^{\theta_{i\beta}})^{n_{i\cdot}}$, consisting of $n_{i\cdot}$ identical factors $x_\beta e^{\theta_{i\beta}}$, one such factor corresponding to each of the $n_{i\cdot}$ elements in the row. To the j-th column of the array corresponds the x_j which appears in each of the factors of ϕ. To the ij-th cell of the array corresponds $e^{\theta_{ij}}$ which appears only in the factors $(x_\beta e^{\theta_{i\beta}})^{n_{i\cdot}}$, and in each of these only as the coefficient of x_j.

The expansion of ϕ consists of all products which can be formed by taking as factors one and only one element $x_\beta e^{\theta_{\alpha\beta}}$ (not summed) from each factor of ϕ. But taking $x_j e^{\theta_{ij}}$ from one of the factors $(x_\beta e^{\theta_{i\beta}})^{n_{i\cdot}}$ of ϕ corresponds exactly to putting an element in the ij-th cell of a lattice such as (27). Hence every term in the expansion of ϕ corresponds to a particular distribution in such a lattice. Moreover, all terms of ϕ correspond to distributions in which the row totals are $n_{\alpha\cdot}$, for we must take $n_{i\cdot}$ elements from the product $(x_\beta e^{\theta_{i\beta}})^{n_{i\cdot}}$. Further, those terms in which the x_β appear in the particular product $x_1^{n_{\cdot 1}} x_2^{n_{\cdot 2}} \cdots x_s^{n_{\cdot s}}$ correspond to distributions in which the column totals are $n_{\cdot 1}, n_{\cdot 2}, \cdots, n_{\cdot s}$, since choosing $n_{\cdot j}$ elements $x_j e^{\theta_{\alpha j}}$ corresponds to putting $n_{\cdot j}$ elements in the j-th column and some row of the lattice.

Expanding ϕ we obtain

$$(29) \qquad \phi = \cdots + \left[\sum \prod_{\alpha=1}^{r} \begin{bmatrix} n_{\alpha\cdot} \\ n_{\alpha\beta} \end{bmatrix} e^{\sum_{\alpha,\beta} n_{\alpha\beta}\theta_{\alpha\beta}}\right] x_1^{n_{\cdot 1}} x_2^{n_{\cdot 2}} \cdots x_s^{n_{\cdot s}} + \cdots$$

where the summation is over all partitions $(n_{\alpha 1} n_{\alpha 2} \cdots n_{\alpha s})$ of the $n_{\alpha\cdot}$ such that $(n_{1\beta} n_{2\beta} \cdots n_{r\beta})$ is also a partition of $n_{\cdot\beta}$. It is clear that since every set of values of the $n_{\alpha\beta}$ subject to the partition restrictions $\sum_\beta n_{\alpha\beta} = n_{\alpha\cdot}$, $\sum_\alpha n_{\alpha\beta} = n_{\cdot\beta}$ corresponds to a particular distribution of n elements in the lattice (27), every particular product $\prod_{\alpha=1}^{r} \begin{bmatrix} n_{\alpha\cdot} \\ n_{\alpha\beta} \end{bmatrix}$ corresponds to such a distribution, and represents the number of ways in which it can arise. Further, the total coefficient displayed (29), namely $\sum \prod_{\alpha=1}^{r} \begin{bmatrix} n_{\alpha\cdot} \\ n_{\alpha\beta} \end{bmatrix}$, represents the total number of ways in which distributions with row totals $n_{\alpha\cdot}$ and column totals $n_{\cdot\beta}$ can arise. Setting all the $\theta_{\alpha\beta} = 0$ we readily find

$$(30) \qquad \begin{aligned} \sum \prod_{\alpha=1}^{r} \begin{bmatrix} n_{\alpha\cdot} \\ n_{\alpha\beta} \end{bmatrix} &= K_{n_{\cdot 1} n_{\cdot 2} \cdots n_{\cdot s}} \phi \big|_{\theta_{\alpha\beta}=0} = K_{n_{\cdot 1} n_{\cdot 2} \cdots n_{\cdot s}} (x_1 + x_2 + \cdots + x_s)^n \\ &= \begin{bmatrix} n \\ n_{\cdot\beta} \end{bmatrix}. \end{aligned}$$

Hence the probability of any particular distribution $\| n_{\alpha\beta} \|$ with fixed row totals $n_{\alpha\cdot}$ and fixed column totals $n_{\cdot\beta}$ is

$$(31) \qquad P(\| n_{\alpha\beta} \| \cdot | n_{\alpha\cdot}, n_{\cdot\beta}) = \frac{\prod_\alpha \begin{bmatrix} n_{\alpha\cdot} \\ n_{\alpha\beta} \end{bmatrix}}{\begin{bmatrix} n \\ n_{\cdot\beta} \end{bmatrix}}.$$

Moments of the n_{ij}. Consider now the result of differentiating ϕ with respect to a particular $\theta_{\alpha\beta}$, say θ_{ij}. We obtain

$$(32) \qquad \frac{\partial\phi}{\partial\theta_{ij}} = \cdots + \sum\nolimits_a n_{ij} \prod_\alpha \begin{bmatrix} n_{\alpha\cdot} \\ n_{\alpha\beta} \end{bmatrix} e^{\sum_{\alpha,\beta} n_{\alpha\beta}\theta_{\alpha\beta}} x_1^{n_{\cdot 1}} x_2^{n_{\cdot 2}} \cdots x_s^{n_{\cdot s}} + \cdots$$

where $\sum_\alpha$ denotes summation over indices such that $\sum_\alpha n_{\alpha\beta} = n_{.\beta}$, $\sum_{\alpha \neq i} n_{\alpha j} + n_{ij} = n_{.j}$ $(\beta \neq j)$. Now $n_{ij} \leq \min(n_{i.}, n_{.j})$, but also n_{ij} can never be less than $n_{.j} - (n - n_{i.})$. For $n_{.j} = n_{ij} + \sum_{\alpha \neq i} n_{\alpha j}$. Since the maximum value of $n_{\alpha j} \leq n_{\alpha.}$, the maximum value of $\sum_{\alpha \neq i} n_{\alpha j} \leq \sum_{\alpha \neq i} n_{\alpha.}$. Hence

$$n_{ij} = n_{.j} - \sum_{\alpha \neq i} n_{\alpha j} \geq n_{.j} - \sum_{\alpha \neq i} n_{\alpha.} = n_{.j} - (n - n_{i.}).$$

Therefore

$$\max(0, n_{.j} - n + n_{i.}) \leq n_{ij} \leq \min(n_{i.}, n_{.j}).$$

Accordingly, combining all the terms of (32) in which n_{ij} has a particular value, γ, we have

$$\frac{\partial \phi}{\partial \theta_{ij}} = \cdots + \sum_{\gamma = \max(0, n_{.j} - n + n_{i.})}^{\min(n_{i.}, n_{.j})} \gamma \sum^*_{\sum_\alpha n_{\alpha\beta} = n_{.\beta}} \prod_\alpha^* \begin{bmatrix} n_{\alpha.} \\ n_{\alpha\beta} \end{bmatrix} \cdot e^{\sum_{\alpha,\beta} n_{\alpha\beta}\theta_{\alpha\beta}} x_1^{n_{.1}} x_2^{n_{.2}} \cdots x_s^{n_{.s}} + \cdots. \tag{33}$$

where Σ^* denotes summation and Π^* multiplication with $n_{ij} = \gamma$.

Since $\sum^*_{\sum_\alpha n_{\alpha\beta} = n_{.\beta}} \prod_\alpha^* \begin{bmatrix} n_{\alpha.} \\ n_{\alpha\beta} \end{bmatrix}$ is precisely the number of distributions $\| n_{\alpha\beta} \|$ for which $n_{ij} = \gamma$, it follows that

$$E(n_{ij} \mid n_{\alpha.}, n_{.\beta}) = \frac{1}{\begin{bmatrix} n \\ n_{.\beta} \end{bmatrix}} K_{n_{.1} n_{.2} \cdots n_{.s}} \left. \frac{\partial \phi}{\partial \theta_{ij}} \right|_{\theta_{\alpha\beta} = 0}. \tag{34}$$

Similarly it follows that

$$E(n_{ij}^p \mid n_{\alpha.}, n_{.\beta}) = \frac{1}{\begin{bmatrix} n \\ n_{.\beta} \end{bmatrix}} K_{n_{.1} n_{.2} \cdots n_{.s}} \left. \frac{\partial^p \phi}{\partial \theta_{ij}^p} \right|_{\theta_{\alpha\beta} = 0} \tag{35}$$

$$E(n_{ij}^p n_{kl}^q \mid n_{\alpha.}, n_{.\beta}) = \frac{1}{\begin{bmatrix} n \\ n_{.\beta} \end{bmatrix}} K_{n_{.1} n_{.2} \cdots n_{.s}} \left. \frac{\partial^{p+q} \phi}{\partial \theta_{ij}^p \, \partial \theta_{kl}^q} \right|_{\theta_{\alpha\beta} = 0} \tag{36}$$

where we may have $i = k$ or $i \neq k$, and $j = l$ or $j \neq l$.

By straightforward differentiation and reduction we find that for the array (27) with given marginal totals $n_{\alpha.}$, $n_{.\beta}$

$$E(n_{ij}) = \frac{n_{i.} n_{.j}}{n} \tag{37}$$

$$E(n_{ij}^2) = \frac{n_{i.}^{(2)} n_{.j}^{(2)}}{n^{(2)}} + \frac{n_{i.} n_{.j}}{n} \tag{38}$$

$$\sigma^2_{n_{ij}} = \frac{[n^2 - n(n_{i.} + n_{.j}) + n_{i.}n_{.j}]n_{i.}n_{.j}}{n^2(n-1)} \tag{39}$$

$$E(n_{ij}^3) = \frac{n_{i.}^{(3)} n_{.j}^{(3)}}{n^{(3)}} + 3\frac{n_{i.}^{(2)} n_{.j}^{(2)}}{n^{(2)}} + \frac{n_{i.}n_{.j}}{n} \tag{40}$$

$$E(n_{ij}^4) = \frac{n_{i.}^{(4)} n_{j.}^{(4)}}{n^{(4)}} + 6\frac{n_{i.}^{(3)} n_{.j}^{(3)}}{n^{(3)}} + 7\frac{n_{i.}^{(2)} n_{.j}^{(2)}}{n^{(2)}} + \frac{n_{i.}n_{.j}}{n}, \tag{41}$$

and if i and k, j and l are distinct

$$E(n_{ij}^2 n_{kj}^2) = \frac{n_{i.}^{(2)} n_{k.}^{(2)} n_{.j}^{(4)}}{n^{(4)}} + (n_{i.}^{(2)} n_{k.} + n_{i.} n_{k.}^{(2)})\frac{n_{.j}^{(3)}}{n^{(3)}} + \frac{n_{i.}n_{k.}n_{.j}^{(2)}}{n^{(2)}} \tag{42}$$

$$E(n_{ij}^2 n_{il}^2) = \frac{n_{i.}^{(4)} n_{.j}^{(2)} n_{.l}^{(2)}}{n^{(4)}} + (n_{.j}^{(2)} n_{.l} + n_{.j} n_{.l}^{(2)})\frac{n_{i.}^{(3)}}{n^{(3)}} + \frac{n_{i.}^{(2)} n_{.j} n_{.l}}{n^{(2)}} \tag{43}$$

$$\begin{aligned} E(n_{ij}^2 n_{kl}^2) = {} & \frac{n_{i.}^{(2)} n_{k.}^{(2)} n_{.j}^{(2)} n_{.l}^{(2)}}{n^{(4)}} + \frac{n_{i.}^{(2)} n_{k.} n_{.j}^{(2)} n_{.l}}{n^{(3)}} \\ & + \frac{n_{i.} n_{k.}^{(2)} n_{.j} n_{.l}^{(2)}}{n^{(3)}} + \frac{n_{i.} n_{k.} n_{.j} n_{.l}}{n^{(2)}} \end{aligned} \tag{44}$$

Moments of the distribution of Chi Square. For the array (27)

$$\begin{aligned} \chi^2 &= \sum_{\alpha,\beta} \frac{\left(n_{\alpha\beta} - \frac{n_{\alpha.}n_{.\beta}}{n}\right)^2}{\frac{n_{\alpha.}n_{.\beta}}{n}} \\ &= \sum_{\alpha,\beta}\left[\frac{n}{n_{\alpha.}n_{.\beta}} n_{\alpha\beta}^2 - 2n_{\alpha\beta} + \frac{n_{\alpha.}n_{.\beta}}{n}\right]. \end{aligned} \tag{45}$$

Hence, using the above results we can, theoretically, find all the moments of the exact distribution of χ^2. It is not difficult to show that

$$E(\chi^2) = \frac{n}{n-1}(r-1)(s-1). \tag{46}$$

The value of $E[(\chi^2)^2]$ and the variance of χ^2 were found by straightforward application of our methods and the results agreed with those given by Haldane [10].

The writer is indebted to Professor Wilks for helpful criticisms and suggestions.

REFERENCES

[1] M. S. Bartlett, "Properties of sufficiency and statistical tests," *Proc. Roy. Soc.*, A, Vol. 160 (1937), pp. 268–282.

[2] Dwight Chapman, "The statistics of the method of correct matchings," *Amer. Jour. Psych.*, Vol. 46 (1934), pp. 287–298.

[3] Dwight Chapman, "The generalized problem of correct matchings," *Annals of Math. Stat.*, Vol. 6 (1935), pp. 85–95.

[4] J. A. Greenwood, "Variance of a general matching problem," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 56–59.

[5] J. A. Greenwood, "Variance of the ESP call series," *Jour. of Parapsychology*, Vol. 2 (1938), pp. 60–64.

[6] J. A. Greenwood, "The first four moments of a general matching problem," *Annals of Eug.*, Vol. 10 (1940), pp. 290–292.

[7] T. N. E. Greville, "Exact probabilities for the matching hypothesis," *Jour. of Parapsychology*, Vol. 2 (1938), pp. 55–59.

[8] T. N. E. Greville, "The frequency distribution of a general matching problem," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 350–354.

[9] J. B. S. Haldane, "The mean and variance of Chi square, when used as a test of homogeneity, when expectations are small," *Biometrika*, Vol. 31 (1940).

[10] J. B. S. Haldane, "The first six moments of Chi-square for an n-fold table with n degrees of freedom when some expectations are small," *Biometrika*, Vol. 29 (1939), p. 389.

[11] J. B. S. Haldane, "The exact value of the moments of the distribution of chi-square, used as a test of goodness of fit, when the expectations are small," *Biometrika*, Vol. 29 (1939), p. 133.

[12] E. V. Huntington, "A rating table for card matching experiments," *Jour. of Parapsychology*, Vol. 4 (1937), pp. 292–294.

[13] E. V. Huntington, "Exact probabilities in certain card-matching problems," *Science*, Vol. 86 (1937), pp. 499–500.

[14] Solomon Kullback, "Note on a matching problem," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 77–80.

[15] E. G. Olds, "A moment-generating function which is useful in solving certain matching problems," *Bull. Amer. Math. Soc.*, Vol. 44 (1938), pp. 407–413.

[16] T. E. Sterne, "The solution of a problem in probability," *Science*, Vol. 86, (1937) pp. 500–501.

[17] W. L. Stevens, "Distribution of entries in a contingency table," *Annals of Eugenics*, Vol. 8 (1938), pp. 238–244.

[18] W. L. Stevens, "Tests of significance for extra sensory perception data," Psychological Review, Vol. 46 (1938), pp. 142–150.

[19] S. S. Wilks, "Statistical aspects of experiments in telepathy," a lecture delivered to the Galois Institute of Mathematics, Long Island University, December 4, 1937.

ON THE CHOICE OF THE NUMBER OF CLASS INTERVALS IN THE APPLICATION OF THE CHI SQUARE TEST

By H. B. Mann and A. Wald[1]

Columbia University

Introduction. To test whether a sample has been drawn from a population with a specified probability distribution, the range of the variable is divided into a number of class intervals and the statistic,

$$\sum_{i=1}^{i=k} \frac{(\alpha_i - Np_i)^2}{Np_i} = \chi^2, \tag{1}$$

computed. In (1) k is the number of class intervals, α_i the number of observations in the ith class, p_i the probability that an observation falls into the ith class (calculated under the hypothesis to be tested). It is known that under the null hypothesis (hypothesis to be tested) the statistic (1) has asymptotically the chi-square distribution with $k - 1$ degrees of freedom, when each Np_i is large. To test the null hypothesis the upper tail of the chi-square distribution is used as a critical region.

In the literature only rules of thumb are found as to the choice of the number and lengths of the class intervals. It is the purpose of this paper to formulate principles for this choice and to determine the number and lengths of the class intervals according to these principles.

If a choice is made as to the number of class intervals it is always possible to find alternative hypotheses with class probabilities equal to the class probabilities under the null hypothesis. The least upper bound of the "distances" of such alternative distributions from the null hypothesis distribution can evidently be minimized by making the class probabilities under the null hypothesis equal to each other. By the distance of two distribution functions we mean the least upper bound of the absolute value of the difference of the two cumulative distribution functions. We have therefore based this paper on a procedure by which the lengths of the class intervals are determined so that the probability of each class under the null hypothesis is equal to $1/k$ where k is the number of class intervals.[2]

Let $C(\Delta)$ be the class of alternative distributions with a distance $\geqslant \Delta$ from the null hypothesis. Let $f(N, k, \Delta)$ be the greatest lower bound of the power of the chi-square test with sample size N and number of class intervals k with respect to alternatives in $C(\Delta)$. The maximum of $f(N, k, \Delta)$ with respect to k is a function $\Phi(N, \Delta)$ of N and Δ. It is most desirable to maximize $f(N, k, \Delta)$ for

[1] Research under a grant in aid from the Carnegie Corporation of New York.

[2] This procedure was first used by H. Hotelling. "The consistency and ultimate distribution of optimum statistics," *Trans. Am. Math. Soc.*, Vol. 32, pp. 851.) It has been advocated by E. J. Gumbel in a paper which will appear shortly.

such values of Δ for which $\Phi(N, \Delta)$ is neither too large nor too small and in this paper we propose to determine Δ so that $\Phi(N, \Delta)$ is equal to $\frac{1}{2}$.

Hence we introduce the following definitions:

DEFINITION 1. *A positive integer k is called best with respect to the number of observations N if there exists a Δ such that $f(N, k, \Delta) = \frac{1}{2}$ and $f(N, k', \Delta) \leqslant \frac{1}{2}$ for any positive integer k'.*

DEFINITION 2. *A positive integer k is called ϵ-best ($0 \leqslant \epsilon \leqslant 1$) with respect to the number of observations N if ϵ is the smallest number in the interval [0, 1] for which the following condition is fulfilled: There exists a Δ such that $f(N, k, \Delta) \geqslant \frac{1}{2} - \epsilon$ and $f(N, k', \Delta) \leqslant \frac{1}{2} + \epsilon$ for any positive integer k'.*

It is obvious that an ϵ-best k is a best k if $\epsilon = 0$. If ϵ is very small an ϵ-best k is for all practical purposes equivalent to a best k.

Since $f(N, k, \Delta)$ is a continuous function of Δ it is easy to see that for any pair of positive integers k and N there exists exactly one value ϵ such that k is ϵ-best with respect to the number of observations N. Since the value of this ϵ is a function of k and N we will denote it by $\epsilon(k, N)$.

DEFINITION 3. *A sequence $\{k_N\}$ of positive integers is called best in the limit if* $\lim\limits_{N=\infty} \epsilon(k_N, N) = 0$.

In this paper the following theorem is proved:

THEOREM 1. *Let* $k_N = 4\sqrt[5]{\dfrac{2(N-1)^2}{c^2}}$ *where c is determined so that* $\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_c^\infty e^{-x^2/2}\,dx$ *is equal to the size of the critical region (probability of the critical region under the null hypothesis) then the sequence $\{k_N\}$ is best in the limit. Furthermore* $\lim\limits_{N=\infty} f(N, k_N, \Delta_N) = \frac{1}{2}$ *for* $\Delta_N = \dfrac{5}{k_N} - \dfrac{4}{k_N^2}$.

It is further shown that for $N \geqslant 450$, if the 5% level of significance is used, and for $N \geqslant 300$, if the 1% level of significance is used, the value of $\epsilon(k_N, N)$ is small so that for practical purposes k_N can be considered as a best k. The authors are convinced although no rigorous proof has been given that $\epsilon(k_N, N)$ is quite small for $N \geqslant 200$ and is very likely to be small even for considerably lower values of N.

1. Mean value and standard deviation of the statistic under alternative hypotheses. It is well known that every continuous distribution can by a simple transformation be transformed into a rectangular distribution with range [0, 1]. We may therefore for convenience assume that the hypothesis to be tested is that of a rectangular distribution with the range [0, 1]. Moreover as mentioned earlier we assume that a procedure is chosen by which the class probabilities under the null hypothesis are equal to each other.

The statistic whose mean value and standard deviation is to be determined is

$$\sum_{i=1}^{i=k} x_i^2 = \chi'^2 \quad \text{where} \quad x_i = \sqrt{\frac{k}{N}}\left(\alpha_i - \frac{N}{k}\right).$$

Let p_i be the probability under the alternative hypothesis that one observation will fall into the ith class. The probability of obtaining certain specified values $\alpha_1, \alpha_2, \cdots, \alpha_k$ is given by

$$f(\alpha_1, \alpha_2, \cdots \alpha_k) = \frac{N!}{\alpha_1! \alpha_2! \cdots \alpha_k!} p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}.$$

Since $\sum_{i=1}^{i=k} \alpha_i = N$ we have

$$\sum_{i=1}^{i=k} x_i^2 = \frac{k}{N} \sum_{i=1}^{i=k} \alpha_i^2 - N.$$

We consider the function

$$(p_1 e^{t_1} + p_2 e^{t_2} + \cdots p_k e^{t_k})^N = \Sigma f(\alpha_1, \alpha_2, \cdots \alpha_k) e^{\alpha_1 t_1 + \alpha_2 t_2 + \cdots \alpha_k t_k}.$$

Differentiating twice and then setting $t_i = 0$ for $i = 1, 2, \cdots k$ we obtain

$$(2) \qquad N(N-1)p_i^2 + Np_i = E(\alpha_i^2), \quad N(N-1)p_i p_j = E(\alpha_i \alpha_j) \text{ for } i \neq j.$$

Hence

$$E\left(\sum_{i=1}^{i=k} \alpha_i^2\right) = N(N-1) \sum_{i=1}^{i=k} p_i^2 + N,$$

and

$$(3) \qquad E(\chi'^2) = k(N-1) \sum_{i=1}^{i=k} p_i^2 + k - N.$$

To compute the standard deviation of χ'^2 we put

$$\mu_i = \left(Np_i - \frac{N}{k}\right)\sqrt{\frac{k}{N}} = \sqrt{Nk}\left(p_i - \frac{1}{k}\right),$$

$$y_i = (\alpha_i - Np_i)\sqrt{\frac{k}{N}} \qquad \text{hence} \qquad y_i = x_i - \mu_i, \qquad E(y_i) = 0.$$

We have

$$\sigma_{\chi'^2}^2 = E\left[\sum_{i=1}^{i=k} (y_i + \mu_i)^2 - E\left(\sum_{i=1}^{i=k} (y_i + \mu_i)^2\right)\right]^2$$

$$= E\left(\sum_{i=1}^{i=k} y_i^2 + 2\sum_{i=1}^{i=k} y_i \mu_i - E\left(\sum_{i=1}^{i=k} y_i^2\right)\right)^2.$$

Let

$$\sqrt{\frac{N}{k}}\, y_i = z_i, \qquad \sqrt{\frac{N}{k}}\, \mu_i = \nu_i;$$

then

$$\nu_i = N\left(p_i - \frac{1}{k}\right), \qquad z_i = \alpha_i - Np_i.$$

We now assume that N is so large that the joint distribution of the z_i is sufficiently well approximated by a multivariate normal distribution. Then

$$E(z_i^2 z_j) = 0, \quad E(z_i^4) = 3[E(z_i^2)]^2, \quad E(z_i^2 z_j^2) = E(z_i^2)E(z_j^2) + 2[E(z_i z_j)]^2 \text{ for } i \neq j.$$

We have the well known relations

$$E(z_i^2) = E(\alpha_i^2) - N^2 p_i^2 = Np_i(1 - p_i),$$
$$E(z_i z_j) = E(\alpha_i \alpha_j) - N^2 p_i p_j = -Np_i p_j .$$

Using the above equations we obtain

$$\sigma^2_{\chi'^2} = \frac{k^2}{N^2}\left\{E\left(\sum_{i=1}^{i=k} z_i^2\right)^2 - \left(E\sum_{i=1}^{i=k} z_i^2\right)^2 + 4E\left(\sum_{i=1}^{i=k} z_i \nu_i\right)^2\right\},$$

$$E\left(\sum_{i=1}^{i=k} z_i^2\right)^2 - \left(E\sum_{i=1}^{i=k} z_i^2\right)^2$$
$$= N^2\left\{3\sum_{i=1}^{i=k} p_i^2(1-p_i)^2 + \sum_{i\neq j}[p_i p_j(1-p_i)(1-p_j) + 2p_i^2 p_j^2] - \left[\sum_{i=1}^{i=k} p_i(1-p_i)\right]^2\right\}$$
$$= 2N^2\left[\sum_{i=1}^{i=k} p_i^2(1-p_i)^2 + \sum_{i\neq j} p_i^2 p_j^2\right]$$
$$= 2N^2\left[\sum_{i=1}^{i=k} p_i^2 - 2\sum_{i=1}^{i=k} p_i^3 + \left(\sum_{i=1}^{i=k} p_i^2\right)^2\right].$$

Further

$$E\left(\sum_{i=1}^{i=k} z_i \nu_i\right)^2 = E\left(\sum_{i=1}^{i=k} z_i^2 \nu_i^2\right) + E\left(\sum_{i\neq j} z_i z_j \nu_i \nu_j\right)$$
$$= N^3\left[\sum_{i=1}^{i=k} p_i(1-p_i)\left(p_i - \frac{1}{k}\right)^2 - \sum_{i\neq j} p_i p_j\left(p_i - \frac{1}{k}\right)\left(p_j - \frac{1}{k}\right)\right]$$
$$= N^3\left[\sum_{i=1}^{i=k} p_i\left(p_i - \frac{1}{k}\right)^2 - \left[\sum_{i=1}^{i=k} p_i\left(p_i - \frac{1}{k}\right)\right]^2\right]$$
$$= N^3\left[\sum_{i=1}^{i=k} p_i^3 - \frac{2}{k}\sum_{i=1}^{i=k} p_i^2 + \frac{1}{k^2} - \left[\sum_{i=1}^{i=k} p_i^2 - \frac{1}{k}\right]^2\right]$$
$$= N^3\left[\sum_{i=1}^{i=k} p_i^3 - \left(\sum p_i^2\right)^2\right].$$

Substituting this into the formula for $\sigma^2_{\chi'^2}$ we finally obtain

$$\sigma^2_{\chi'^2} = 2k^2\left\{\sum_{i=1}^{i=k} p_i^2 + 2(N-1)\sum_{i=1}^{i=k} p_i^3 - (2N-1)\left(\sum_{i=1}^{i=k} p_i^2\right)^2\right\}. \tag{4}$$

2. The Taylor expansion of the power. Let C be determined so that the probability under the null hypothesis that $\sum_{i=1}^{i=k} x_i^2 \geqslant C$ is equal to the size λ_0 of

the critical region. Let $P\left(\sum_{i=1}^{i=k} x_i^2 \geqslant C\right)$ be the probability under the alternative hypothesis that $\sum_{i=1}^{i=k} x_i^2 \geqslant C$. Then the power P is given by

$$P\left(\sum_{i=1}^{i=k} x_i^2 \geqslant C\right), \tag{5}$$

where

$$x_i = \frac{\alpha_i - \frac{N}{k}}{\sqrt{\frac{N}{k}}}.$$

Hence

$$\sum_{i=1}^{i=k} x_i^2 = \frac{k}{N}\left(\sum_{i=1}^{i=k} \alpha_i^2 - \frac{N^2}{k}\right),$$

and (5) can be written in the form

$$P\left(\sum_{i=1}^{i=k} \alpha_i^2 \geqslant C'\right) \tag{6}$$

where C' is a certain function of N and k. Let $\delta_i = p_i - \frac{1}{k}$, where p_i is the probability of the ith class interval under the alternative hypothesis.

Expanding P into a power series we obtain (in this and the following derivations, we take all partial differential quotients at the point $\delta_1 = \delta_2 = \cdots = \delta_k = 0$)

$$P = \lambda_0 + \sum_{i=1}^{i=k} \delta_i \frac{\partial P}{\partial \delta_i} + \frac{1}{2}\left\{\sum_{i=1}^{i=k} \delta_i^2 \frac{\partial^2 P}{\partial \delta_i^2} + \sum_{i \neq j} \delta_i \delta_j \frac{\partial^2 P}{\partial \delta_i \partial \delta_j}\right\} + \cdots.$$

Since P is a symmetric function of the δ_i we have for $\delta_1 = \delta_2 = \cdots = \delta_k = 0$

$$\frac{\partial^2 P}{\partial \delta_i^2} = \frac{\partial^2 P}{\partial \delta_1^2}, \qquad \frac{\partial^2 P}{\partial \delta_i \partial \delta_j} = \frac{\partial^2 P}{\partial \delta_1 \partial \delta_2} \qquad \text{for } i \neq j.$$

Furthermore $\sum_{i=1}^{i=k} \delta_i = 0$. Therefore

$$P = \lambda_0 + \frac{1}{2}\left\{\frac{\partial^2 P}{\partial \delta_1^2} \sum_{i=1}^{i=k} \delta_i^2 + \frac{\partial^2 P}{\partial \delta_1 \partial \delta_2} \sum_{i \neq j} \delta_i \delta_j\right\} + \cdots.$$

We shall first show that the terms of second order are always positive. This shows that the test is unbiased and justifies again the choice of equal class probabilities under the null hypothesis since this assures unbiasedness and mini-

mizes among all unbiased tests the g.l.b. of the distances of such alternatives whose power is equal to the size of the critical region.
The power is given by

$$P = \sum_{\alpha_1^2+\alpha_2^2+\cdots\alpha_k^2 \geq C'} \frac{N!}{\alpha_1!\alpha_2!\cdots\alpha_k!} p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}.$$

Since $\sum_{i=1}^{i=k} \delta_i^2 = -\sum_{i\neq j} \delta_i\delta_j$ we obtain for the second order terms

$$\frac{\partial^2 P}{\partial \delta_1^2} \sum_{i=1}^{i=k} \delta_i^2 + \frac{\partial^2 P}{\partial\delta_1\partial\delta_2} \sum_{i\neq j} \delta_i\delta_j = \left(\frac{\partial^2 P}{\partial\delta_1^2} - \frac{\partial^2 P}{\partial\delta_1\partial\delta_2}\right)\sum_{i=1}^{i=k}\delta_i^2$$

$$(7) \qquad = \sum_{\alpha_1^2+\alpha_2^2+\cdots+\alpha_k^2\geq C'} (\alpha_1^2 - \alpha_1 - \alpha_1\alpha_2)p(\alpha_1, \alpha_2 \cdots \alpha_k) \sum_{i=1}^{i=k} \delta_i^2$$

where

$$p(\alpha_1, \cdots \alpha_k) = \frac{N!}{\alpha_1!\alpha_2!\cdots\alpha_k!}\frac{1}{k^N}.$$

In the following derivation extend all sums if not otherwise stated over all terms for which $\sum_{i=1}^{i=k} \alpha_i^2 \geqslant C'$ and use the relation $\sum_{i=1}^{i=k} \alpha_i = N$. We have because of the symmetry

$$\sum \alpha_1 p(\alpha_1, \alpha_2, \cdots \alpha_k) = \frac{N}{k}\sum p(\alpha_1, \alpha_2 \cdots \alpha_k) = \frac{N}{k}\lambda_0,$$

$$\sum \alpha_1\alpha_2 p(\alpha_1, \alpha_2, \cdots \alpha_k) = \frac{1}{k(k-1)}\sum\left(N^2 - \sum_{i=1}^{i=k}\alpha_i^2\right)p(\alpha_1, \alpha_2, \cdots \alpha_k)$$

$$= \frac{N^2\lambda_0}{k(k-1)} - \frac{1}{k-1}\sum \alpha_1^2 p(\alpha_1, \alpha_2, \cdots \alpha_k).$$

Hence the coefficient of the second order term becomes

$$\frac{k}{k-1}\sum \alpha_1^2 p(\alpha_1, \alpha_2, \cdots \alpha_k) - \frac{N}{k}\lambda_0 - \frac{N^2}{k(k-1)}\lambda_0$$

$$= \frac{1}{k-1}\sum\sum_{i=1}^{i=k}\alpha_i^2 p(\alpha_1, \alpha_2, \cdots \alpha_k) - \frac{N}{k}\lambda_0 - \frac{N^2}{k(k-1)}\lambda_0.$$

But

$$\frac{\sum\sum_{i=1}^{i=k}\alpha_i^2 p(\alpha_1, \alpha_2, \cdots \alpha_k)}{\lambda_0} > E\left(\sum_{i=1}^{i=k}\alpha_i^2\right),$$

since the conditional mean for values of $\sum_{i=1}^{i=k}\alpha_i^2 \geqslant C'$ must be larger than the

mean of all values of $\sum_{i=1}^{i=k} \alpha_i^2$. Since $E\left(\sum_{i=1}^{i=k} \alpha_i^2\right) = \frac{N^2}{k} - \frac{N}{k} + N$, we obtain

$$\frac{1}{k-1} \sum \sum_{i=1}^{i=k} \alpha_i^2 p(\alpha_1, \alpha_2 \cdots \alpha_k)$$

$$> \frac{\lambda_0}{k-1}\left(\frac{N^2}{k} + \frac{N(k-1)}{k}\right) = \lambda_0\left(\frac{N^2}{k(k-1)} + \frac{N}{k}\right)$$

and hence the coefficient of $\sum_{i=1}^{i=k} \delta_i^2$ is larger than 0.

To prove Theorem 1, we will have to determine the alternative distribution for which $\sum_{i=1}^{i=k} \delta_i^2$ becomes a minimum subject to the condition that the distance from the null hypothesis should be greater than or equal to a given Δ.

Hence we have to find a distribution function $F(x)$ such that $| F(x) - x | \geqslant \Delta$ for at least one value x and $\sum_{i=1}^{i=k} \delta_i^2 = \sum_{i=1}^{i=k} \left(p_i - \frac{1}{k}\right)^2 = \sum_{i=1}^{i=k} p_i^2 - \frac{1}{k}$ is a minimum where $p_i = F\left(\frac{i}{k}\right) - F\left(\frac{i-1}{k}\right)$. Instead of minimizing $\sum_{i=1}^{i=k} \delta_i^2$ we may minimize $\sum_{i=1}^{i=k} p_i^2$, since the two expressions differ merely by a constant. There will be two different solutions for $F(x)$ depending on whether $F(x) - x \geqslant \Delta$ or $F(x) - x \leqslant -\Delta$ for at least one value x. Because of symmetry we restrict ourselves to the case in which $F(x) - x \geqslant \Delta$ for at least one value of x.

Let a be a value for which $F(a) - a \geqslant \Delta$ and suppose that

$$\frac{l-1}{k} < a \leqslant \frac{l}{k}$$

then

$$F(a) \geqslant a + \Delta,$$

$$F\left(\frac{l}{k}\right) = \frac{l}{k} + \epsilon.$$

We prove first

$$\epsilon \geqslant \Delta - \frac{1}{k}.$$

Proof: Since $F\left(\frac{l}{k}\right) - F(a) \geqslant 0$ we have

$$F\left(\frac{l}{k}\right) = F(a) + F\left(\frac{l}{k}\right) - F(a) \geqslant a + \Delta$$

and

$$\epsilon = F\left(\frac{l}{k}\right) - \frac{l}{k} \geqslant a + \Delta - \frac{l}{k} \geqslant \frac{l-1}{k} + \Delta - \frac{l}{k} = \Delta - \frac{1}{k}.$$

If $\Delta \leqslant \frac{1}{k}$ we can always find a distribution function in $C(\Delta)$ for which $p_i = \frac{1}{k}$. Hence we consider only the case $k > \frac{1}{\Delta}$. We must minimize $\sum_{i=1}^{i=k} p_i^2$ under the condition $\sum_{i=1}^{i=l} p_i = \frac{l}{k} + \epsilon, \sum_{i=l+1}^{i=k} p_i = \frac{k-l}{k} - \epsilon$. We therefore minimize

$$\Phi = \sum_{i=1}^{i=k} p_i^2 - 2\lambda_1 \sum_{i=1}^{i=l} p_i - 2\lambda_2 \sum_{i=l+1}^{i=k} p_i .$$

This leads to

$$p_i = \begin{cases} \dfrac{1}{k} + \dfrac{\epsilon}{l} & \text{for} \quad i = 1, \cdots l \\[2ex] \dfrac{1}{k} - \dfrac{\epsilon}{k-l} & \text{for} \quad i = (l+1), \cdots k. \end{cases}$$

We then have

$$\sum_{i=1}^{i=k} p_i^2 = l\left(\frac{1}{k} + \frac{\epsilon}{l}\right)^2 + (k-l)\left(\frac{1}{k} - \frac{\epsilon}{k-l}\right)^2 = \frac{1}{k} + \frac{\epsilon^2 k}{l(k-l)} .$$

This is smallest if $\epsilon = \Delta - \frac{1}{k}$ and $l = \frac{k}{2}$. The following discontinuous distribution function gives these values for ϵ, l and p_i and has the distance Δ from the rectangular distribution.

$$\begin{aligned} &F(x) = x\left[1 + 2\left(\Delta - \frac{1}{k}\right)\right] && \text{for} \quad 0 \leqslant x \leqslant \frac{1}{2} - \frac{1}{k}, \\ &F(x) = \frac{1}{2} + \Delta - \frac{1}{k} && \text{for} \quad \frac{1}{2} - \frac{1}{k} < x \leqslant \frac{1}{2}, \\ (8) \quad &F(x) = x\left[1 - 2\left(\Delta - \frac{1}{k}\right)\right] + 2\left(\Delta - \frac{1}{k}\right) && \text{for} \quad \frac{1}{2} \leqslant x \leqslant 1, \\ &F(x) = 0 && \text{for} \quad 0 \leqslant x, \\ &F(x) = 1 && \text{for} \quad x \geqslant 1. \end{aligned}$$

3. Solution for large N. Denote by $F(\Delta, k)$ the distribution function (8) of $C(\Delta)$ which makes $\sum_{i=1}^{i=k} \delta_i^2$ a minimum if the test is made with k class intervals.

Assume that k is large enough that χ'^2 can be taken as normally distributed. The power of the test is then given by

$$\frac{1}{\sqrt{2\pi}\,\sigma'}\int_{(k-1)+c\sqrt{2(k-1)}}^{\infty} e^{-\frac{1}{2\sigma'^2}\left(\sum_{i=1}^{i=k} x_i^2 - E\left(\sum_{i=1}^{i=k} x_i^2\right)\right)^2} d\left(\sum_{i=1}^{i=k} x_i^2\right) \tag{9}$$
$$= \frac{1}{\sqrt{2\pi}}\int_{\frac{k-1-E\left(\sum_{i=1}^{i=k} x_i^2\right)+c\sqrt{2(k-1)}}{\sigma'}}^{\infty} e^{-\frac{1}{2}y^2}\,dy,$$

where σ' is the standard deviation of $\sum_{i=1}^{i=k} x_i^2$ and c is determined so that $\frac{1}{\sqrt{2\pi}}\int_c^\infty e^{-\frac{1}{2}y^2}\,dy$ is equal to the size of the critical region. Hence to maximize the power with respect to k is equivalent to maximizing

$$\psi(k) = \frac{E\left(\sum_{i=1}^{i=k} x_i^2\right) - (k-1) - c\sqrt{2(k-1)}}{\sigma'}$$

with respect to k.

Under the alternative $F(\Delta, k)$ we obtain

$$E\left(\sum_{i=1}^{i=k} x_i^2\right) - (k-1) = k(N-1)\sum_{i=1}^{i=k} p_i^2 + k - N - k + 1 = 4(N-1)\left(\Delta - \frac{1}{k}\right)^2$$

Hence

$$\psi(k) = \frac{4(N-1)\left(\Delta - \frac{1}{k}\right)^2 - c\sqrt{2(k-1)}}{\sigma'}.$$

We choose Δ so that this maximum power is exactly $\frac{1}{2}$, that is, so that $\psi(k) = 0$ for that k which maximizes $\psi(k)$. Denote this value of Δ by Δ_N and let k_N be the value of k which maximizes $\psi(k)$. The differential-quotient of the numerator of $\psi(k)$ with respect to k is then equal to 0 for $k = k_N$. Hence

$$8(N-1)\left(\Delta_N - \frac{1}{k_N}\right)\frac{1}{k_N^2} = \frac{c}{\sqrt{2(k_N-1)}}. \tag{10}$$

Furthermore since $\psi(k_N) = 0$ we have

$$4(N-1)\left(\Delta_N - \frac{1}{k_N}\right)^2 = c\sqrt{2(k_N-1)}. \tag{11}$$

Solving equations (10) and (11) we obtain

$$\Delta_N = \frac{5}{k_N} - \frac{4}{k_N^2} \tag{12}$$

and

$$\sqrt[5]{\frac{k_N^8}{(k_N-1)^5}} = 4\sqrt[5]{\frac{2(N-1)^2}{c^2}}$$

or since $k_N > 3$,

$$k_N < 4\sqrt[5]{\frac{2(N-1)^2}{c^2}} < k_N + 1.$$

Hence

$$\text{(13)} \quad \text{either } k_N = \left[4\sqrt[5]{\frac{2(N-1)^2}{c^2}}\right] \quad \text{or} \quad k_N = \left[4\sqrt[5]{\frac{2(N-1)^2}{c^2}}\right] + 1,$$

is the value of k for which the power with respect to $F(\Delta_N, k)$ becomes a maximum. We have merely to show that $\psi''(k)$ is negative for $k = k_N$.

Using the fact that $\psi(k_N) = \psi'(k_N) = 0$ we obtain

$$\sigma'\psi''(k_N) = \frac{-16(N-1)}{k_N^3}\Delta_N + \frac{24(N-1)}{k_N^4} + \frac{c}{(\sqrt{2(k_N-1)})^3}.$$

Substituting for Δ_N the right hand side of (12) we obtain on account of (10)

$$\sigma'\psi''(k_N) = \frac{-56(N-1)}{k_N^4} + \frac{64(N-1)}{k_N^5} + \frac{8(N-1)}{2(k-1)}\left(\frac{4}{k_N^3} - \frac{4}{k_N^4}\right).$$

Using $2(k-1) > k$ we obtain

$$\psi''(k_N) < \frac{1}{k^4\sigma'}\left(-24(N-1) + \frac{32}{k}(N-1)\right)$$

which is negative. σ' can be shown to be of order $k_N^{\frac{1}{2}}$; $\psi''(k_N)$ is, therefore, of order $\frac{N}{k_N^{4+\frac{1}{2}}} = 0\left(\frac{1}{N^{\frac{4}{5}}}\right)$. The maximum is, therefore, rather flat for large values of N.

We shall now show that if k is large enough to assume χ'^2 to be normally distributed then $F(\Delta, k)$ is the alternative which gives the smallest power compared with all alternatives in the class $C(\Delta)$ provided the power for the alternative $F(\Delta, k)$ equals $\frac{1}{2}$.

We know that $E\left(\sum_{i=1}^{i=k} x_i^2\right)$ is smallest for $F(\Delta, k)$. Since the power with respect to $F(\Delta, k)$ equals $\frac{1}{2}$ we have

$$E\left(\sum_{i=1}^{i=k} x_i^2\right) - (k - 1 - c\sqrt{2(k-1)}) = 0.$$

Thus the lower limit of the integral in (9) becomes negative for every other alternative and the power will be larger than $\frac{1}{2}$.

The power with respect to $F(\Delta_N, k_N)$ is equal to $\frac{1}{2}$, hence if we choose $k = k_N$ the power of the test will be $\geqslant \frac{1}{2}$ for all alternatives in the class $C(\Delta_N)$. On the other hand if we choose $k \neq k_N$ then there will be at least one alternative in

[3] Cantelli's formula and its proof are given by Fréchet in his book *Recherches Théoriques Modernes sur la Théorie de Probabilités*, Paris (1937), pp. 123–126.

$C(\Delta_N)$ for which the power is $<\frac{1}{2}$. (For instance $F(\Delta_N, k)$ is such an alternative.)

The above statements have been derived under the assumption that χ'^2 is normally distributed. Hence if the distribution of χ'^2 were exactly normal $k_N = 4\sqrt[5]{\frac{2(N-1)^2}{c^2}}$ would be a best k and for this k_N and $\Delta_N = \frac{5}{k_N} - \frac{4}{k_N^2}$ the greatest lower bound of the power in the class $C(\Delta_N)$ would be exactly $\frac{1}{2}$. Since the distribution of χ'^2 approaches the normal distribution with $k \to \infty$ the sequence $\{k_N\}$ is best in the limit and Theorem 1 stated in the introduction is proved.

For the purposes of practical applications, it is not enough to know that $\{k_N\}$ is best in the limit. We have to know for what values of N k_N can be considered practically as a best k, i.e. for what values of N the quantity $\epsilon(k_N, N)$ defined in the introduction is sufficiently small. The quantity $\epsilon(k_N, N)$ is certainly small if for the number of class intervals k_N the distribution of χ'^2 is near to normal and if the power with respect to at least one alternative of the class $C(\Delta_N)$ is smaller than $\frac{1}{2}$ also in the case when the number of class intervals is too small to assume a normal distribution for χ'^2.

We shall in the following assume that for $k > 13$ the normal distribution is a sufficiently good approximation. Actually we need not assume a normal distribution but only that the probability is close to $\frac{1}{2}$ that the statistic will exceed its mean value.

Cantelli[3] gave the following formula. Let M_r be the rth moment of a distribution about x_0. Let d be any arbitrary positive number. Let $P(|x - x_0| \leqslant d)$ be the probability that $|x - x_0| \leqslant d$ then the following inequalities hold:

$$\text{If} \quad \frac{M_r}{d^r} \leqslant \frac{M_{2r}}{d^{2r}} \quad \text{then} \quad P(|x - x_0| \leqslant d) \geqslant 1 - \frac{M_r}{d^r}.$$

$$\text{If} \quad \frac{M_r}{d^r} \geqslant \frac{M_{2r}}{d^{2r}} \quad \text{then} \quad P(|x - x_0| \leqslant d) \geqslant 1 - \frac{M_{2r} - M_r^2}{(d^r - M_r)^2 + M_{2r} - M_r^2}.$$

Since χ'^2 can only take positive values we have

$$\text{(14)} \quad \text{If} \quad \frac{E(\chi'^2)}{c_k} \leqslant \frac{\sigma^2_{\chi'^2} + [E(\chi'^2)]^2}{c_k^2} \quad \text{then} \quad P(\chi'^2 \leqslant c_k) \geqslant 1 - \frac{E(\chi'^2)}{c_k}.$$

$$\text{(15)} \quad \text{If} \quad \frac{E(\chi'^2)}{c_k} \geqslant \frac{\sigma^2_{\chi'^2} + [E(\chi'^2)]^2}{c_k^2}$$

$$\text{then} \quad P(\chi'^2 \leqslant c_k) \geqslant 1 - \frac{\sigma^2_{\chi'^2}}{(c_k - E(\chi'^2))^2 + \sigma^2_{\chi'^2}}.$$

Where c_k is determined so that $P(\chi^2 \geqslant c_k)$ equals the size of the critical region if the null hypothesis is true and the number of class intervals equals k. c_k can be obtained from a table of the chi-square distribution.

For $F(\Delta_N, k)$ we obtain with $\Delta'_N = \frac{5}{k_N} - \frac{4}{k_N^2} - \frac{1}{k}$ from (3) and (4)

$$E(\chi'^2) = (k - 1) + 4(N - 1)\Delta_N'^2,$$
$$\sigma^2_{\chi'^2} = 2(k - 1) + 8\Delta_N'^2(k + 2N - 4) - 32(2N - 1)\Delta_N'^4.$$

By numerically calculating $E(\chi'^2)$ and $\sigma_{\chi'^2}$ for $N = 450$ and a 5% level of significance, for $N = 300$ and a 1% level of significance, and for $k = 13, 12 \cdots \left[\frac{1}{\Delta_N}\right] + 1$ it can be shown that for these values of N and k

$$\frac{E(\chi'^2)}{c_k} \geqslant \frac{\sigma^2_{\chi'^2} + [E(\chi'^2)]^2}{c_k^2}. \tag{16}$$

Hence we have to use (15). From (16) it follows that $c_k > E(\chi'^2)$. If $P(\chi'^2 \leqslant c_k \leqslant \frac{1}{2}$ we obtain on account of (15) and (16)

$$\frac{\sigma^2_{\chi'^2}}{(c_k - E(\chi'^2))^2 + \sigma^2_{\chi'^2}} \geqslant \frac{1}{2}, \qquad \sigma_{\chi'^2} + E(\chi'^2) \geqslant c_k.$$

Numerical calculation shows that for the values of N and k and the significance levels considered

$$\sigma_{\chi'^2} + E(\chi'^2) < c_k. \tag{17}$$

It can then be shown that for $N \geqslant 450$ and $N \geqslant 300$ respectively $N\Delta_N'$ decreases with N. A simple argument then shows that (16) and (17) are also true for all values $N \geqslant 450$ and $N \geqslant 300$ respectively. Hence the power with respect to $F(\Delta_N, k)$ is $<\frac{1}{2}$ for these values of N. Thus we see: For $N \geqslant 450$ if the 5% level is used, and for $N \geqslant 300$ if the 1% level is used, the value $k_N = 4\sqrt[5]{\frac{2(N-1)^2}{c^2}}$ can be considered for practical purposes as a best k. The value c is determined so that $\frac{1}{\sqrt{2\pi}}\int_c^\infty e^{-\frac{1}{2}t^2}\,dt$ is equal to the size of the critical region.

LIMITED TYPE OF PRIMARY PROBABILITY DISTRIBUTION APPLIED TO ANNUAL MAXIMUM FLOOD FLOWS

By Bradford F. Kimball

Port Washington, N. Y.

1. Theoretical statement of problem. There is no doubt that Gumbel's recent paper "The Return Period of Flood Flows" [1] has supplied an admirably simple technique for engineers to use in approximating the trend of *return periods* of annual maximum flood flows for purposes of extrapolation. This treatment is scientifically of great interest because it introduces for the first time into a subject already treated at considerable length by engineers, the theory of the probability distribution of maximum values as developed by Fisher and Tippett, von Mises, and others.[1] However, certain further observations should be made concerning the approach used by Gumbel.

Let x represent the measure of daily stream flow having a probability distribution $w(x)$. Let the probability distribution of the associated annual maximum stream flows be denoted by $V(x)$ with

$$W(x) = \int_0^x V(s)\, ds, \tag{1}$$

denoting probability that annual maxima be less than or equal to x. The *return period* $T(x)$ of an annual maximum flow of measure x is then defined by

$$T(x) = \frac{1}{1 - W(x)}. \tag{2}$$

In this paper the probability distribution $w(x)$ will be called the *primary* probability distribution associated with the probability distribution of maximum values $V(x)$ and its *cumulative* distribution $W(x)$.

Gumbel argues that for the type of primary probability distribution that might reasonably be expected to apply, $W(x)$ will be of the type introduced by R. A. Fisher:

$$W(x) = \exp\,[-\exp - \alpha(x - u)]. \tag{3}$$

It is further implied that a primary probability distribution involving an upper limit would lead to a probability distribution of maximum values of the type

$$W_1(x) = \frac{k}{u}\left(\frac{u}{x}\right)^{k+1} \cdot e^{-(u/x)^k}, \tag{4}$$

for which moments of order k or higher do not exist. The inference is then drawn that a primary probability distribution leading to such a cumulative distribution of maximum values would seem to be less likely to be the correct

[1] See references at end of Gumbel's paper, loc. cit.

one than one leading to the distribution (3). To this argument we do not object; but we question the implied conclusion that *hence the use of a limited type of primary distribution is to be disallowed.*

If the primary probability distribution be of the *limited* Galton type

$$w(x) = K \exp\left(-\tfrac{1}{2}u^2\right), \tag{5}$$

where K is a constant and

$$u = k[b - \log(a - x)], \qquad 0 \leqq x \leqq a, \tag{6}$$

it can be shown that the limiting form of the cumulative distribution of maxima of n values takes the same type form (3) where x is replaced by u. This can be seen by observing that the transformed variate u becomes infinite as x approaches a, and hence has infinite range to the right, which places (5) in the category of distributions which are known to lead to cumulative distribution of maximum values of form (3). More explicitly, considering $w(x)$ as a finite distribution in x, if one traces the reasoning as set forth in von Mises' derivation [2] of the limiting distribution (3), one finds that since the cumulative primary probability $\int_0^x w(s)\,ds$ does *not* have a non-vanishing derivative of finite order at $x = a$, that what von Mises terms *the case of a limited distribution* does *not* apply, while the argument for a cumulative distribution of maxima of form (3) *does* carry through, in spite of the fact that x has limited range to the right. This fact was not mentioned by Gumbel.

One is thus led to the conclusion that there is no logical exclusion of the assumption of a primary probability distribution of the form (5).

One might well argue for a first approximation of the actual primary probability distribution of stream flows—using any regular time interval such as a day or an hour—of the form (5). Differentiating u with respect to x, one obtains

$$k\,dx = (a - x)\,du, \tag{7}$$

which means that to a constant probability increment Δu there corresponds a maximum increment Δx in measure of stream flow equal to $(a/k)\Delta u$ when x is at the lower limit zero. This corresponding increment in stream flow decreases linearly to zero as x approaches its upper bound a, imposed because of the existence of a finite watershed.

2. Technique of fitting probability distribution of maximum values in case primary probability distribution is of the limited type (5)–(6). Write the cumulative maximum distribution (3) in the form:

$$\begin{aligned} W(x) &= \exp(-\exp -y), \qquad y = \alpha(u(x) - u_1), \\ u(x) &= k[b - \log(a - x)], \qquad 0 \leqq x \leqq a. \end{aligned} \tag{8}$$

Now it is known that for the distribution

$$dW = e^{-e^{-y}} e^{-y} \, dy, \tag{9}$$

the mean value and standard deviation of y are given by

$$\begin{aligned} \bar{y} &= .577215 \text{ (Euler's constant } C) \\ \sigma^2(y) &= \pi^2/6. \end{aligned} \tag{10}$$

Hence

$$\bar{y} = \alpha[\bar{u}(x) - u_1] = \alpha k[(b - u_1/k) - \bar{L}] = C,$$

where $\bar{L}$ denotes the mean value of $\log (a - x)$, with x representing the observed *maximum flood flows*. Also

$$\sigma(y) = \alpha k \, \sigma(L) = \pi/\sqrt{6}$$

where $\sigma(L)$ denotes the standard deviation of $\log (a - x)$. Hence

$$\alpha k = (\pi/\sqrt{6})/\sigma(L), \qquad b - u_1/k = C/\alpha k + \bar{L}, \tag{11}$$

and y is determined as a function of x by the relation

$$y = \alpha k[(b - u_1/k) - \log (a - x)]. \tag{12}$$

It is interesting to observe that it has not been necessary to determine the constants k and b of the primary probability distribution. Only the upper bound a and observed flood flows are used in this process. From the relation (12) the theoretical curve in terms of x may easily be computed from tables relating y to W (See Gumbel, loc. cit., Table II, page 173).

The difficulty of determining what the upper bound a should be in a specific case is a practical one and does not concern the objective theoretical problem of choosing the *type* of curve which most nearly describes the behavior of annual maximum flood flows. The point to be made in this paper is that the use of what seems to be a reasonable value of a, *will* materially alter forecasts of future annual flood flows relative to forecasts made on the assumption that such an upper limit may be neglected. It is also ventured that the resulting theoretical probability distribution of maxima will in general give a better fit to the series of observed floods than one based on the latter premise. Techniques for determination of upper bound a will not be discussed in this paper.

3. Examples. In order to demonstrate the point in question the two methods have been applied to a 57 year record of the annual flood flows of the Tennessee River at Chattanooga for the years 1875 to 1931.[2]

---

[2] The author has already used this series in a previous article [3] and for this reason has found it convenient to use it here.

TABLE I

Series of observed annual flood flows

(Tennessee River at Chattanooga, 1875–1931)

(1) Observed Flood x	(2) Ratio to Mean	(3) Per cent of Time	(4) Return Period, $T(x)$
85.9	.412	0.88	1.007
108	.518	2.63	1.027
123	.590	4.39	1.043
			.....
310	1.487	95.61	22.8
349	1.674	97.37	38.0
361	1.731	99.12	114.

In Table I, col. (1) is shown the incomplete series of observed annual floods in units of 1,000 c.f.s. arranged in order of magnitude. The complete series may be referred to in *Water-Supply Paper 771* entitled "Floods in the United States," *U. S. Geological Survey*, 1936, p. 401. The mean annual maximum flood of this series is 208.56. The ratio of each annual maximum to the mean is shown in Col. (2). In Col. (4) is shown the observed return period which is taken here as the harmonic mean between what has been called the *exceedance interval* and the *recurrence interval* (see Gumbel, loc. cit., Table I, p. 167). Thinking of the 57 year record as a span of 57 years, the above procedure is equivalent to taking the observed probability $W(x)$ that a given annual flood will not be exceeded as the mid-point of the part of this time-span covered by the observed flood in question. Thus the lowest flood-peak 85,900 c.f.s. corresponds to the span from zero to 1.754 per cent of the whole time-span, and hence $W(x)$ is taken at the mid-point, -0.877 per cent. Similarly the greatest flood, 361,000 c.f.s. corresponds to interval from 98.246 to 100 per cent and is taken at 99.12 per cent. These arithmetic means correspond to harmonic means of the "recurrence" and "exceedance" intervals referred to above. This is the procedure which Hazen [4] originally followed.

Data from Cols. (1) and (4) of this table determined position of dots on Fig. 1. Data from Cols. (2) and (3) gave the points indicated by dots on Fig. 2, with $1 - W(x)$ recorded on the chart rather than $W(x)$.

The two theoretical distributions fitted to these annual flood maxima will be referred to as distributions A and B.

Distribution A. In this case the limited type of primary probability distribution (5) – (6) is assumed. From previous studies of this data series made by the author [3], an upper limit of annual floods of some 609,000 c.f.s. was found to be reasonable, and for purposes of this example the same upper limit will be assumed for the primary probability distribution. Thus the transformation (6) becomes:

$$u = k[b - \log (609 - x)],$$

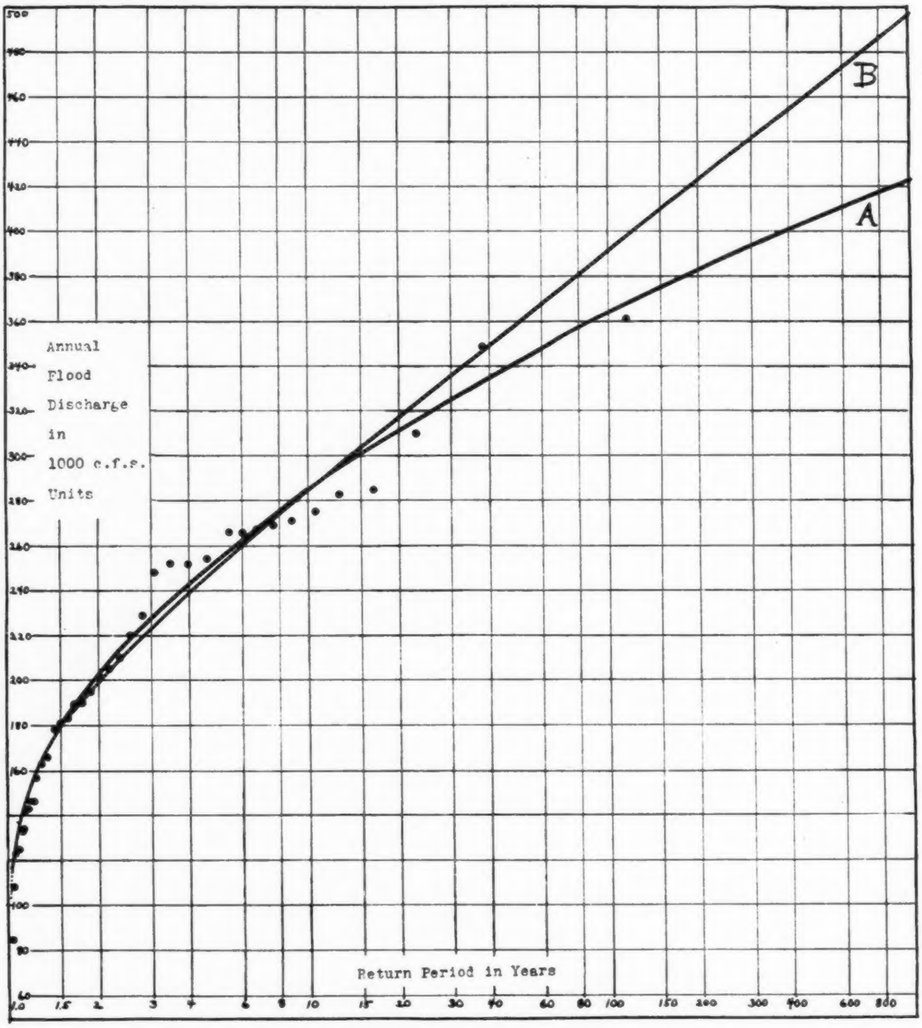

FIG. 1. Comparison of methods of fitting annual flood peaks, (Tennessee River at Chattanooga, 1875–1931)—return periods plotted against annual flood discharges on semi-logarithmic chart.

where the logarithm to base 10 can be used without loss of generality since the constant k will absorb the conversion factor. The mean value of the logarithm, and its standard deviation come to

$$\bar{L} = 2.59772, \qquad \sigma(L) = .06576$$

The constants of the transformation (12) are thus determined by

$$\alpha k = (\pi/\sqrt{6})/(.06576), \qquad b - u_1/k = C/(\alpha k) + 2.59772$$

Thus

$$1/(\alpha k) = .05127, \qquad b - u_1/k = 2.6273$$

and solving (12) for log $(609 - x)$,

$$\log (609 - x) = 2.6273 - (.05127)\, y \tag{13}$$

Using a table for the known relations between y, $W(x)$, and $T(x)$ for the Fisher-Tippett distribution of maximum values similar to Table II of Gumbel's article

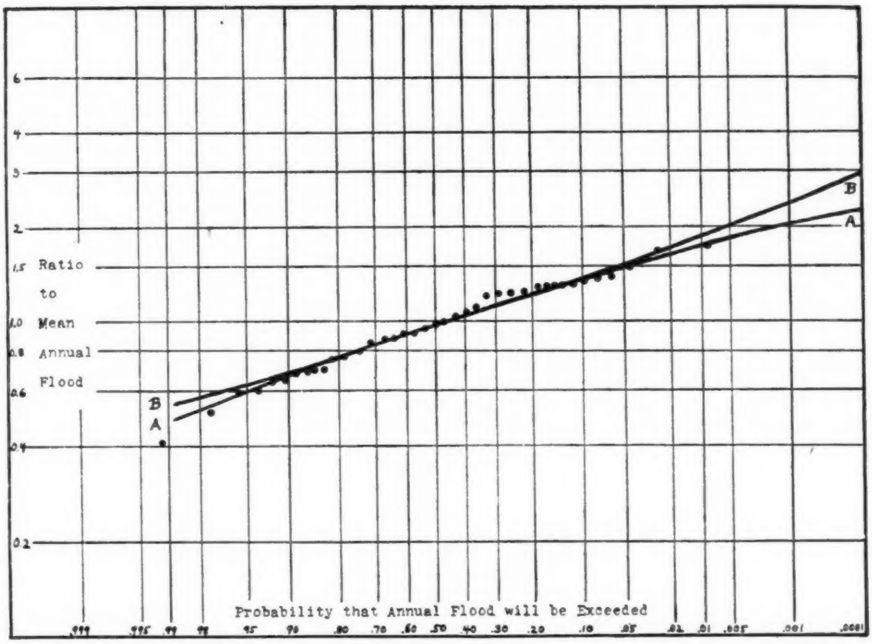

FIG. 2. Comparison of methods of fitting annual flood peaks, (Tennessee River at Chattanooga, 1875–1931)—Data plotted on logarithmic probability chart designed by Hazen, Whipple and Fuller.

(loc. cit.) the corresponding values x of the annual floods are easily determined. Thus a theoretical relation between x and $W(x)$ is set up. This is indicated as Curve A on the two charts exhibited here.

Distribution B. The primary probability distribution in this case is taken as unlimited to the right, and in general is assumed to have the character of an exponentially decreasing function of the measure of stream flow x (see Gumbel, loc. cit.). The parameter y of the distribution of annual maxima is given directly by

$$y = \alpha(x - x_1)$$

and

$$1/\alpha = (\sqrt{6}/\pi) \text{ (stand. dev. of annual floods)} = (.77970)\ (58.26) = 45.425$$

$$x_1 = \text{(mean annual flood)} - C/\alpha = 208.6 - (.57722)\ (45.425) = 182.4$$

Hence

$$x = 182.4 - (45.425)\, y \tag{14}$$

and using the table of corresponding values of y, $W(x)$ and $T(x)$ for the Fisher-Tippett distribution referred to above, a theoretical relation between x and $W(x)$ is easily set up. This is plotted as Curve B on the accompanying charts.

4. Discussion of examples. In Fig. 1 it is to be noted that if theoretical curves are continued to the right to give readings for a return period of 1,000 years, the divergence of Curve A from Curve B is large enough to be of significance, numerically. Visual inspection does not indicate which curve is the better fit to the observation points.

In Fig. 2 the curves are plotted on "logarithmic probability" graph paper. This paper was designed by Hazen and Fuller [4] specifically for the purpose of plotting annual maxima of stream-flows. A significant divergence in trend is to be noted at the right hand end.

These charts indicate that the use of an upper limit may materially affect extrapolation of fitted theoretical curves, for purposes of estimating floods with a return period, say of 1,000 years.

If the trends of observed floods in Gumbel's recent paper in the *Transactions of the American Geophysical Union* [5] are examined, it will be observed that in the case of the Connecticut, Mississippi and Rhone rivers, there is a decided tendency for the curve of observed floods to turn downwards, away from the theoretical curves, which correspond to Curve B exhibited in Figure 1. In the case of the Tennessee, Cumberland and Columbia rivers the tendency is not decisive, while in the case of the Rhine river at Basel (Switzerland) the tendency of the observed curve is upwards rather than downwards. As the writer has observed elsewhere [6], this last data series seems to be rather unique in character and is possibly the result of a watershed greatly influenced by all year around snow deposits. Possibly a radically different primary probability distribution should be used in this case.

5. Conclusion. The writer has demonstrated in this paper that in fitting a theoretical probability distribution of maximum values to annual maxima of stream flows, the use of an upper bound for measures of stream flow by assumption of a primary probability distribution of the type (5)–(6)

(1) is not inconsistent with the use of the Fisher-Tippett distribution of maxima,

(2) has a reasonable logical basis from the point of view of the hydrologist,

(3) may materially affect the estimation of return periods when extrapolation is involved, relative to results obtained when no upper bound is assumed.

It has not been within the scope of this paper to discuss techniques for determining such an upper bound, nor to apply the theory to enough data series to draw conclusions concerning goodness of fit.

REFERENCES

[1] E. J. GUMBEL, "The return period of flood flows," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 163–190.

[2] RICHARD VON MISES, "La distribution de la plus grande de *n* valeurs," *Revue Mathematique de l'Union Interbalkanique*, Vol. 1, Athens (1936).

[3] BRADFORD F. KIMBALL, "Probability-distribution-curve for flood-control studies," *Trans. Am. Geographical Union*, 1938, pp. 466–475.

[4] ALLEN HAZEN, *Flood-flows*, John Wiley & Sons, New York, 1930.

[5] E. J. GUMBEL, "Probability-interpretation of the observed return-periods of floods," *Trans. Am. Geophysical Union*, 1941, pp. 836–850.

[6] BRADFORD F. KIMBALL, *Discussion* of paper entitled "Statistical control curves for flood discharges" by E. J. Gumbel, *Trans. Am. Geophysical Union*, 1942.

LINEAR RESTRICTIONS ON CHI-SQUARE

By Franklin E. Satterthwaite

University of Iowa

Chi-square is a statistic widely used in statistical analysis. It is usually of the form,

$$
\begin{aligned}
\chi^2 &= \textstyle\sum_1^n \chi_i^2 \\
&= \textstyle\sum_1^n \left(\dfrac{x_i - m_i}{\sigma_i}\right)^2,
\end{aligned} \tag{1}
$$

where the x_i's are independent normally distributed variables drawn from populations with respective means and standard deviations, m_i and σ_i. In practical problems the independence of the x_i's is often modified by placing restrictions on the χ_i's in order to estimate the m_i's or σ_i's. It is well known that if m such restrictions which are linear and homogeneous (also algebraically independent) are placed on the χ_i's, then the resulting chi-square, (1), is distributed according to the chi-square distribution with $n - m$ degrees of freedom. The purpose of this paper is to study the case where the restrictions are not necessarily homogeneous.

1. Geometrical development. The χ_i's of equation (1) may be considered as co-ordinates in an n-dimensional space. Equation (1) represents a sphere in such a space with its center at the origin and with radius, χ. We should like to determine the distribution of χ^2. First, since the χ_i's are independent, we may form their joint distribution,[1]

$$
\begin{aligned}
F(\chi_1, \chi_2, \cdots \chi_n)\, dV &= K\Pi_j e^{-\frac{1}{2}\chi_j^2}\, d\chi_j \\
&= Ke^{-\frac{1}{2}\Sigma\chi_j^2}\, d\chi_1\, d\chi_2 \cdots d\chi_n \\
&= Ke^{-\frac{1}{2}\chi^2}\, dV.
\end{aligned} \tag{2}
$$

We may change the variable in (2) to χ^2 if we can determine dV. Since the n-dimensional sphere represented by equation (1) has a volume proportional to χ^n, we may write

$$
\begin{aligned}
dV &= Kd(\chi^2)^{\frac{1}{2}n} \\
&= K(\chi^2)^{\frac{1}{2}n-1}\, d\chi^2.
\end{aligned}
$$

Substituting this value in the distribution (2) we obtain for the distribution of chi-square,

$$
F(\chi^2)\, d\chi^2 = K(\chi^2)^{\frac{1}{2}n-1} e^{-\frac{1}{2}\chi^2}\, d\chi^2,
$$

which is the usual form of the chi-square distribution for n degrees of freedom.

[1] The letter K will be used throughout as a constant, not necessarily the same constant from equation to equation.

We shall next restrict the values of χ_i by means of a condition,

$$(3) \qquad a_{11}\chi_1 + a_{12}\chi_2 + \cdots + a_{1n}\chi_n = \rho_1, \qquad \Sigma a_{1j}^2 = 1,$$

where ρ_1 is a constant. This restriction represents a hyper-plane in our n-dimensional space at a distance ρ_1 from the origin. The intersection of this hyper-plane with our sphere (1) is an $(n - 1)$-dimensional sphere with radius

$$\chi' = (\chi^2 - \rho_1^2)^{\frac{1}{2}}.$$

The differential of the volume of this sphere is

$$dV = K(\chi^2 - \rho_1^2)^{\frac{1}{2}(n-1)-1} d\chi^2.$$

Substituting this in the distribution (2) we obtain the distribution of chi-square subject to the single linear restriction, (3). Thus

$$F(\chi^2)\, d\chi^2 = K(\chi^2 - \rho_1^2)^{\frac{1}{2}(n-1)-1} e^{-\frac{1}{2}\chi^2}\, d\chi^2,$$

or more conveniently,

$$F(\chi^2 - \rho_1^2)\, d(\chi^2 - \rho_1^2) = K(\chi^2 - \rho_1^2)^{\frac{1}{2}(n-1)-1} e^{-\frac{1}{2}(\chi^2-\rho_1^2)}\, d(\chi^2 - \rho_1^2).$$

The argument may be readily extended to include additional linear restrictions of the form,

$$(4) \qquad \begin{aligned} a_{21}\chi_1 + a_{22}\chi_2 + \cdots + a_{2n}\chi_n &= \rho_2, \qquad \Sigma a_{2j}^2 = 1, \\ &\cdots \\ a_{m1}\chi_1 + a_{m2}\chi_2 + \cdots + a_{mn}\chi_n &= \rho_m, \qquad \Sigma a_{mj}^2 = 1. \end{aligned}$$

For convenience we shall assume that the restrictions form an orthogonal set[2] so that

$$\Sigma_j a_{ij} a_{kj} = 0, \qquad i \neq k.$$

The hyper-plane represented by equation (4) is at a distance, ρ_2, from the origin. Since (4) is orthogonal to (3), it is also at a distance, ρ_2, from the center of the $(n - 1)$-dimensional sphere obtained on applying the first restriction. Therefore the intersection of this hyper-plane with the $(n - 1)$-dimensional sphere will give an $(n - 2)$-dimensional sphere of radius

$$\chi'' = (\chi^2 - \rho_1^2 - \rho_2^2)^{\frac{1}{2}}.$$

Similarly, if we consider all m restrictions, we obtain an $(n - m)$-dimensional sphere with radius

$$\chi^{(m)} = (\chi^2 - \Sigma\rho_j^2)^{\frac{1}{2}}.$$

[2] Any set of linear restrictions which are algebraically independent and consistent may be replaced by an orthogonal set. Thus if (4) were not orthogonal to (3), we could replace (4) by (4) $- k$(3) where k is determined by the condition

$$\Sigma a_{1j}(a_{2j} - k a_{1j}) = 0$$

or

$$\Sigma a_{1j} a_{2j} = k \Sigma a_{1j}^2$$

The differential of the volume of this sphere will be

$$dV = K(\chi^2 - \Sigma\rho_j^2)^{\frac{1}{2}(n-m)-1}\, d(\chi^2 - \Sigma\rho_j^2).$$

Substituting this in (2) we see that

$$(\chi^{(m)})^2 = \chi^2 - \sum \rho_j^2$$

is distributed as is chi-square with $n - m$ degrees of freedom.

2. Alternate analytic development. It is perhaps desirable that we present an analytic proof of the foregoing theorem. Therefore we shall first regard the ρ_j's as variables and shall determine the joint distribution of χ^2 and the ρ_j's. We may then pass to the distribution of those values of χ^2 which correspond to assigned values of the ρ_j's. Note that the χ_i's are considered to be statistically independent.

The characteristic function of the joint distribution of χ^2 and the ρ_j's is known to be[3]

$$\phi(t, t_1, \cdots t_m) = \frac{e^{-Q/2(1-2it)}}{(1 - 2it)^{\frac{1}{2}n}},$$

where

$$\begin{aligned} Q &= \sum_{i,j,k} a_{ik} a_{jk} t_i t_j \\ &= \sum t_i^2, \qquad\qquad \text{since } \sum a_{ik} a_{jk} = \delta_{ij}. \end{aligned}$$

Applying the Fourier transform, we obtain the joint distribution of χ^2 and the ρ_j's:

$$F(\chi^2, \rho_1, \cdots \rho_n) = K \int \cdots \int \frac{e^{Q'}}{(1 - 2it)^{\frac{1}{2}n}}\, dt_m \cdots dt_1\, dt,$$

where

$$\begin{aligned} Q' &= -it\chi^2 - \Sigma it_j\rho_j - \{\Sigma t_j^2/2(1 - 2it)\} \\ &= -it\chi^2 - \frac{\Sigma[t_j + i\rho_j(1 - 2it)]^2}{2(1 - 2it)} - \tfrac{1}{2}(1 - 2it)\Sigma\rho_j^2. \end{aligned}$$

Performing the integration with respect to $t_1, \cdots t_m$, we have,

$$F = Ke^{-\frac{1}{2}\Sigma\rho_j^2} \int \frac{e^{-it\chi^2}}{(1 - 2it)^{\frac{1}{2}n}} (1 - 2it)^{\frac{1}{2}m} e^{it\Sigma\rho_j^2}\, dt,$$

and finally,

$$F = K(\chi^2 - \Sigma\rho_j^2)^{\frac{1}{2}(n-m)-1} e^{-\frac{1}{2}\chi^2}.$$

[3] See A. T. Craig, "A certain mean value problem in statistics," *Bull. Amer. Math. Soc.*, Vol. 42 (1936), p. 671.

In our problem we want the distribution of χ^2 (or more conveniently, of $\chi^2 - \Sigma\rho_j^2$) when the ρ_j's take on fixed values. To obtain this we substitute fixed values, $\hat{\rho}_j$'s, into the joint distribution and divide by the marginal total,

$$\int F(\chi^2, \hat{\rho}_1, \hat{\rho}_2 \cdots \hat{\rho}_m)\, d\chi^2 = K\Gamma\,[\tfrac{1}{2}(n-m)]2^{\frac{1}{2}(n-m)}\, e^{-\frac{1}{2}\Sigma\hat{\rho}_j^2}.$$

This gives us the distribution function,

$$F(\chi^2 - \Sigma\hat{\rho}) = \frac{1}{2\Gamma[\frac{1}{2}(n-m)]}\,[\tfrac{1}{2}(\chi^2 - \Sigma\hat{\rho}_j^2)]^{\frac{1}{2}(n-m)-1}\, e^{-\frac{1}{2}(\chi^2-\Sigma\hat{\rho}_j^2)},$$

which is a chi-square distribution with $n - m$ degrees of freedom.

3. Application. As an example of the use of linear restrictions on chi-square we shall now examine the effect on the chi-square test of goodness of fit if the moments of a sample are not corrected for grouping errors in fitting a frequency curve.

The parameters of the fitted frequency distribution, $f(x)$, are determined from the equations,

$$N\int x^k f(x)\, dx = \Sigma\, x_j^k \theta_j, \qquad k = 0, 1, 2, \cdots, \tag{5}$$

where x_j is the mid-point of the j^{th} group and θ_j the corresponding observed frequency. Next a set of expected frequencies,

$$\hat{\theta}_j = \int_{\alpha_j}^{\alpha_{j+1}} Nf(x)\, dx, \qquad \alpha_j = (x_{j-1} + x_j)/2,$$

is determined by taking partial areas of the fitted frequency distribution. The expected frequency is used to transform the actual frequency into a statistic with mean zero and unit variance by the equation,

$$\chi_j = (\theta_j - \hat{\theta}_j)/\hat{\theta}^{\frac{1}{2}}.$$

Equations (5) may now be rearranged into the form of linear restrictions on the χ_j. Thus

$$\Sigma\, x_j^k \hat{\theta}_j^{\frac{1}{2}} \chi_j = \rho_k' \tag{6}$$

where the ρ_k' have the values,

$$\begin{aligned}\rho_k' &= \sum x_j^k \theta_j - \Sigma\, x_j^k \hat{\theta}_j \\ &= N\int x^k f(x)\, dx - \Sigma\, x_j^k \hat{\theta}_j \\ &\neq 0 \text{ in general}\end{aligned}$$

To make our example more specific, let us fit a normal distribution to a sample of 1000 items with mean zero and unit variance. Let the grouping be about the midpoints,

x_j : −3, −2, −1, 0, 1, 2, 3.

The expected frequencies in each group are

$\hat{\theta}_j$: 6, 61, 242, 382, 242, 61, 6.

The variance of these expected frequencies is 1.080 as contrasted with 1.000 for the sample. The linear restrictions, (6), now take the forms,

$$(7)\quad 2.4\chi_{-3} + 7.8\chi_{-2} + 15.6\chi_{-1} + 19.5\chi_0 + 15.6\chi_1 + 7.8\chi_2 + 2.4\chi_3 = 0$$

$$(8)\quad -7.2\chi_{-3} - 15.6\chi_{-2} - 15.6\chi_{-1} + 0 + 15.6\chi_1 + 15.6\chi_2 + 7.2\chi_3 = 0$$

$$(9)\quad 21.6\chi_{-3} + 31.2\chi_{-2} + 15.6\chi_{-1} + 0 + 15.6\chi_1 + 31.2\chi_2 + 21.6\chi_3 = -80.$$

Because of the symmetry of the normal distribution, restriction (8) is orthogonal to (7) and (9). Therefore the only orthogonalization necessary is to replace (9) by an equivalent restriction which is orthogonal to (7). This can be done by subtracting 1.080 times (7) from (9) which gives

$$(10)\ 19.0\chi_{-3} + 22.8\chi_{-2} - 1.2\chi_{-1} - 21.1\chi_0 - 1.2\chi_1 + 22.8\chi_2 + 19.0\chi_3 = -80$$

If these restrictions are each divided by the square root of the sum of the squares of the coefficients of the χ_j, they will be the normal orthogonal set required by the development. The distances of these restrictive planes from the center of χ^2-sphere are

$$\rho_{(7)} = 0, \qquad \rho_{(8)} = 0, \qquad \rho_{(10)} = 1.7.$$

Thus if we test the goodness of fit of the normal distribution to this sample by calculating chi-square,

$$\chi^2 = \Sigma\, \chi_j^2 = \Sigma \frac{(\theta_j - \hat{\theta}_j)^2}{\hat{\theta}_j},$$

we should subtract from χ^2 a correction of

$$\Sigma\, \rho_k^2 = 2.8$$

before judging the significance. This correction adjusts for the effect of the grouping error on the chi-square test.

In this example, chi-square has four degrees of freedom so that an error of 2.8 is large enough to affect our judgment of its significance. It can be shown that the correction is proportional to the size of the sample. Therefore, if our sample had contained only 100 items, the fit obtained by ignoring grouping effects would be almost as good as the fit when the sample moments were corrected for grouping. On the other hand, if the sample had 10,000 items, it

would be practically impossible to obtain a satisfactory fit without correcting for grouping errors.

4. Conclusion. The theory of the loss of degrees of freedom for chi-square when the underlying statistics are subject to linear restrictions does not require the restrictions to be homogeneous. For restrictions which are not homogeneous, a correction must be subtracted from chi-square equal to the square of the distance from the center of the sphere,

$$\chi^2 = \Sigma\, \chi_i^2 = 0$$

to the intersection of the restrictive planes. Non-homogeneous restrictions sometimes arise in practice because of the bias introduced by an approximation. An example is given from curve fitting.

SYSTEMS OF LINEAR EQUATIONS WITH COEFFICIENTS SUBJECT TO ERROR

By A. T. Lonseth

Iowa State College

1. Introduction. Various scientific problems lead to non-homogeneous systems of n linear equations in n unknowns, in which the $n^2 + n$ coefficients (including "absolute" terms) are subject to error. Such errors may be errors of observation, or errors introduced by rounding off decimal expansions. If the system has a non-vanishing determinant, the ordinary rules yield the solution. But the question arises: how may the possible errors in the coefficients affect the solutions? In particular, one would like to know how to exclude the fatal event that some malicious combination of errors might make the determinant zero. One would further like to have limitations on the solution-errors in terms of maximum coefficient-errors. Considering the coefficient-errors as random variables, one may also inquire as to the probability distributions of the solution-errors.

The principal result obtained in this paper is the Taylor's expansion of the error in any unknown, considered as a function of the $n(n + 1)$ errors in the coefficients. An upper bound is obtained for each term of this series, and the sum of these upper bounds (when convergent) is expressed in closed form. Thus are obtained not only approximations to the maximum error, but an actual upper limit. Convergence of the power series is established for sufficiently small coefficient-errors; "sufficient smallness" is specified in terms of a simple criterion, which simultaneously provides a sufficient condition for the non-vanishing of a determinant with elements subject to error.

These results were obtained before I learned that work had already been done on the problem. The earliest seems to be that of F. R. Moulton [2] in 1913; he found the first order approximation (6) for $n = 3$, and discussed the geometrical reasons for sensitivity. Much later I. M. H. Etherington [1], evidently unaware of Moulton's paper, found the expression for the total error of a determinant whose elements may be in error, and applied this to the present problem. He thus found limits for the first and second order errors, in a rather different form from mine. The probabilistic considerations of section 5 were suggested by Etherington's article. L. B. Tuckerman [3] recently discussed the question of estimating computational errors incurred in the course of solution. He considered only errors of first order.

My original procedure was to compute the terms of the Taylor's series as successive differentials of the unknown, from Cramer's formula. This soon becomes laborious, and I found only the first two terms. The linear matrix equation (4) was then kindly suggested to me by R. Oldenburger. Here (4) is solved by iteration, resulting in a simple recursion formula for successive terms of the Taylor's series.

2. Formal matrix solution. Let the system of equations be

$$\sum_{j=1}^{n} a_{ij}x_j = c_i \qquad i = 1, 2, \cdots, n. \tag{1}$$

In terms of the matrices

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}, \qquad \mathbf{X} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \qquad \mathbf{C} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix},$$

system (1) can be written

$$\mathbf{AX} = \mathbf{C}. \tag{2}$$

Supposing that not all c's vanish, and that A, the determinant of $\mathbf{A}$, does not vanish, there is a unique solution $\mathbf{X}$. But the a's and c's, and consequently the x's, are subject to error: let the true value of a_{ij} be $a_{ij} + \alpha_{ij}$; of c_i, $c_i + \gamma_i$; and of the resulting x_j, $x_j + \xi_j$. We must actually deal with the system

$$(\mathbf{A} + \mathbf{a})(\mathbf{X} + \mathbf{x}) = \mathbf{C} + \mathbf{c}, \tag{3}$$

where we have written

$$\mathbf{a} = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & & \vdots \\ \alpha_{n1} & \cdots & \alpha_{nn} \end{pmatrix}, \qquad \mathbf{x} = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix}, \qquad \mathbf{c} = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{pmatrix}.$$

Expanding (3) and using (2), we find for the error-matrix $\mathbf{x}$

$$\mathbf{x} = \mathbf{m} + \mathbf{nX} + \mathbf{nx}, \tag{4}$$

with $\mathbf{m} = \mathbf{A}^{-1}\mathbf{c}$, $\mathbf{n} = -\mathbf{A}^{-1}\mathbf{a}$; $\mathbf{A}^{-1}$ is the inverse of $\mathbf{A}$. We solve (4) formally for $\mathbf{x}$ by iteration. Thus

$$\mathbf{x} = \mathbf{m} + \mathbf{nX} + \mathbf{n}(\mathbf{m} + \mathbf{nX}) + \mathbf{n}^2\mathbf{x}, \text{ etc.}:$$

and there results the infinite expansion

$$\mathbf{x} = \sum_{k=1}^{\infty} \mathbf{x}^{(k)}: \mathbf{x}^{(1)} = \mathbf{m} + \mathbf{nX}; \qquad \mathbf{x}^{(k)} = \mathbf{nx}^{(k-1)}, \qquad k > 1. \tag{5}$$

In section 4 convergence of (5) will be established for sufficiently small $| \alpha_{ij} |$.

3. The elements of $\mathbf{x}^{(k)}$. It is necessary to consider closely the individual elements of $\mathbf{x}^{(k)}$. Writing

$$\mathbf{x}^{(k)} = \begin{pmatrix} \xi_1^{(k)} \\ \vdots \\ \xi_n^{(k)} \end{pmatrix},$$

we note from (5) that

$$\xi_i = \sum_{k=1}^{\infty} \xi_i^{(k)} ;$$

this is precisely the Taylor's series for the error in x_j : each $\xi_j^{(k)}$ is a homogeneous polynomial of degree k in the α's and γ's. Writing A_{ij} for the cofactor of a_{ij} in A,

$$\begin{aligned}
\mathbf{x}^{(1)} &= \mathbf{m} + \mathbf{nX} = \mathbf{A}^{-1}(\mathbf{c} - \mathbf{aX}) \\
&= \begin{pmatrix} \frac{A_{11}}{A} & \cdots & \frac{A_{n1}}{A} \\ \vdots & & \vdots \\ \frac{A_{1n}}{A} & \cdots & \frac{A_{nn}}{A} \end{pmatrix} \left\{ \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{pmatrix} - \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & & \vdots \\ \alpha_{n1} & \cdots & \alpha_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right\} \\
&= \begin{pmatrix} \frac{A_{11}}{A} & \cdots & \frac{A_{n1}}{A} \\ \vdots & & \vdots \\ \frac{A_{1n}}{A} & \cdots & \frac{A_{nn}}{A} \end{pmatrix} \begin{pmatrix} \gamma_1 - \alpha_{11} x_1 - \cdots - \alpha_{1n} x_n \\ \vdots \\ \gamma_n - \alpha_{n1} x_1 - \cdots - \alpha_{nn} x_n \end{pmatrix},
\end{aligned}$$

whence (summing hereafter from 1 to n on Greek-letter subscripts)

$$\text{(6)} \qquad \xi_j^{(1)} = \frac{1}{A} \left\{ \sum_\mu \gamma_\mu A_{\mu j} - x_1 \sum_\mu \alpha_{\mu 1} A_{\mu j} - \cdots - x_n \sum_\mu \alpha_{\mu n} A_{\mu j} \right\}.$$

From (5), if $k > 1$,

$$\begin{aligned}
\mathbf{x}^{(k)} &= \mathbf{nx}^{(k-1)} = -\mathbf{A}^{-1}\mathbf{ax}^{(k-1)} \\
&= \begin{pmatrix} -\frac{1}{A} \Sigma \alpha_{\mu 1} A_{\mu 1} & \cdots & -\frac{1}{A} \Sigma \alpha_{\mu n} A_{\mu 1} \\ \vdots & & \vdots \\ -\frac{1}{A} \Sigma \alpha_{\mu 1} A_{\mu n} & \cdots & -\frac{1}{A} \Sigma \alpha_{\mu n} A_{\mu n} \end{pmatrix} \begin{pmatrix} \xi_1^{(k-1)} \\ \vdots \\ \xi_n^{(k-1)} \end{pmatrix},
\end{aligned}$$

so that

$$\text{(7)} \qquad \xi_j^{(k)} = -\frac{1}{A} \sum_\nu \xi_\nu^{(k-1)} \sum_\mu \alpha_{\mu\nu} A_{\mu j}, \qquad k > 1.$$

The sums $\Sigma \gamma_\mu A_{\mu j}$, $\Sigma \alpha_{\mu l} A_{\mu j}$ have obvious interpretations as determinants.

4 Bounds and convergence of the series. Assuming $|\alpha_{ij}|$, $|\gamma_i| \leqq \delta$ and taking absolute values in (6),

$$\text{(8)} \qquad |\xi_j^{(1)}| \leqq \frac{\delta}{|A|} (1 + \sum_\mu |x_\mu|)(\sum_\mu |A_{\mu j}|).$$

It will be observed that equality can be attained for a particular choice of α's and γ's as $\pm\delta$: the bound for first-order errors is best possible. But it is not in general possible by a single choice of α's and γ's to obtain equality for all j.

Similarly from (7)

$$|\xi_j^{(k)}| \leqq \frac{\delta}{|A|}(\sum_\mu |\xi_\mu^{(k-1)}|)(\sum_\mu |A_{\mu j}|), \qquad k > 1;$$

whence by induction

$$|\xi_j^{(k)}| \leqq \left(\frac{\delta}{|A|}\right)^k (1 + \sum_\mu |x_\mu|)(\sum_\nu \sum_\mu |A_{\mu\nu}|)^{k-1}(\sum_\mu |A_{\mu j}|). \tag{9}$$

Summing on k,

$$\sum_{k=1}^m |\xi_j^{(k)}| \leqq \frac{\delta}{|A|}(1 + \sum_\mu |x_\mu|)(\sum_\mu |A_{\mu j}|)\left(\sum_{k=1}^m \rho^{k-1}\right),$$

with

$$\rho = \frac{\delta}{|A|}\sum_\nu \sum_\mu |A_{\mu\nu}|.$$

If $\rho < 1$, we can let $m \to \infty$:

$$|\xi_j| \leqq \frac{\delta}{|A|}(1 + \sum_\mu |x_\mu|)(\sum_\mu |A_{\mu j}|)/(1 - \rho). \tag{10}$$

Observing that the γ's occur linearly in (6) and (7), we conclude that (5) converges if

$$|\alpha_{ij}| \leqq \delta < |A|/(\sum_\nu \sum_\mu |A_{\mu\nu}|). \tag{11}$$

It follows that the determinant of the system (3) cannot vanish if (11) holds. This is rather remarkable, in that $\delta\Sigma\Sigma\ |A_{\mu\nu}|$ is merely the maximum first-order term in the error of that determinant ([1], p. 108); the effect of higher order terms (i.e., of any but first-order minors) in producing a zero determinant can be wholly ignored.

From the remark after (8), it appears that equality in (9) and (10) cannot generally be attained.

If (10) is written $|\xi_j| \leqq B/(1 - \rho)$, it is easily seen that the remainder after the hth approximation does not exceed $\rho^h B/(1 - \rho)$.

5. Probability distributions. We now consider some consequences of the following assumptions: the α's and γ's are identical, independent random variables, bounded by a δ satisfying (11), and distributed symmetrically about zero. (It would be reasonable to assume further that they possess a frequency function, which is nowhere concave upward.) Writing $\mathcal{E}(x)$ for "expectation of the random variable x," we have

$$\mathcal{E}(\alpha_{ij}) = \mathcal{E}(\gamma_i) = 0, \qquad \mathcal{E}(\alpha_{ij}^2) = \mathcal{E}(\gamma_i^2) = \sigma^2 < \delta^2.$$

On account of independence and symmetry, the expectation of any power-product of α's and γ's containing an odd power must be zero. To first order, the mean a_j of the solution-error ξ_j is approximated by

$$(12) \qquad a_j^{(1)} = \mathcal{E}(\xi_j^{(1)}) = 0;$$

and the standard deviation S_j by

$$(13) \qquad S_j^{(1)} = \sqrt{\mathcal{E}\{(\xi_j^{(1)})^2\}} \frac{\sigma}{|A|} \{(1 + \sum_\mu x_\mu^2)(\sum_\mu A_{\mu j}^2)\}^{\frac{1}{2}}.$$

The second approximation to a_j is also easily obtained:

$$(14) \qquad a_j^{(2)} = \mathcal{E}(\xi_j^{(2)}) = \frac{\sigma^2}{A^2}(\sum_\nu \sum_\mu x_\nu A_{\mu\nu} A_{\mu j}).$$

Both (13) and (14) were given by Etherington [1], though in a less symmetric form. Higher approximations, as he remarks, involve complicated summations; but if they should ever be required, the machinery exists in (6) and (7) for their systematic computation. As to the errors in using (13) for the standard deviation S_j and (14) for the mean, we know only that

$$a_j = a_j^{(2)} + o(\delta^4), \qquad S_j^2 = (S_j^{(1)})^2 + o(\delta^4).$$

Etherington ([1], p. 111) considers the important special case of "rounding off" decimal expressions. Each a and c is supposed correct in the qth decimal place, the $(q + 1)$th figure being "forced," i.e., increased by one when the $(q + 2)$th figure is dropped, if the $(q + 2)$th is 5, 6, 7, 8, or 9. Assuming constant frequency 10^{-q} in the interval $(-\frac{1}{2}10^{-q}, \frac{1}{2}10^{-q})$, we may use (13) and (14) with $\sigma^2 = 10^{-2q}/12$.

Errors of observation are often assumed to be normally distributed. There is nothing against such an assumption with regard to the γ's, but the α's must not make (3) singular, and must accordingly be suitably bounded, e.g. by (11).

6. Conclusion. The formulas and bounds of this paper involve only these quantities: the determinant A, its first order minors, and the solutions of (1). They can be found in the course of solving (1) by orthodox methods.

Inequality (10) definitely limits the maximum solution-errors, in terms of the maximum coefficient-error δ, provided δ satisfies (11). But it may be that (8), either alone or in conjunction with the second-order bound from (9), will give a better approximation.

The ratio $\Sigma\Sigma \mid A_{\mu\nu} \mid / \mid A \mid$ may be taken as a "measure of sensitivity" of (1) to error.

The fundamental formulas (6) and (7) are capable of solving other problems than those studied here. For example, it may happen that only certain elements (such as those of a single column) are in error, in which case better inequalities can be found. Or the α's and γ's may not be independently and identically distributed.

REFERENCES

[1] I. M. H. Etherington, "On errors in determinants," *Proc. Edinburgh Math. Soc.*, Ser. 2, Vol. 3 (1932), pp. 107–117.

[2] F. R. Moulton, "On the solutions of linear equations having small determinants," *Amer. Math. Monthly*, Vol. 20 (1913), pp. 242–249.

[3] L. B. Tuckerman, "On the mathematically significant figures in the solution of simultaneous linear equations," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 307–316.

[4] P. G. Hoel, "The errors involved in evaluating correlation determinants," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 58–65.

ON MUTUALLY FAVORABLE EVENTS

By Kai-Lai Chung

Tsing Hua University, Kunming, China

Introduction. For a set of arbitrary events, E. J. Gumbel, M. Fréchet and the author[1] have recently obtained inequalities between sums of certain probability functions. One of the results of the author is the following:

Let $E_1, \cdots, E_n$ be n arbitrary events and let $p_m(\nu_1, \cdots, \nu_k)$ denote the probability of the occurrence of at least m events out of the k events $E_{\nu_1}, \cdots, E_{\nu_k}$. Then, for $k = 1, \cdots, n - 1$ and $1 \leqq m \leqq k$ we have

$$\binom{n-m}{k-m} \Sigma p_m(\nu_1, \cdots, \nu_{k+1}) \leqq \binom{n-m}{k-m+1} \Sigma p_m(\nu_1, \cdots, \nu_k),$$

where the summations extend respectively to all combinations of $k + 1$ and k indices out of the n indices $1, \cdots, n$.

In course of proof of the above inequalities it appears that similar inequalities between products instead of sums can be obtained under certain assumptions regarding the nature of interdependence of the events. We shall first study the nature of such assumptions, and then proceed to the proof of the said inequalities (Theorems 1 and 2). It may be noted that the inductive method used here serves equally well for the proof of the inequalities cited above, though somewhat longer, but apparently our former method is not applicable here.

That events satisfying our assumptions actually exist, is shown by an application to the elementary theory of numbers. The author feels incompetent to discuss other possible fields of application.

1. Let a set of events be given

$$E_1, E_2, \cdots, E_n, \cdots$$

and let E_i' denote the event non-E_i. Let $p(i)$ denote the probability of the occurrence of E_i, $p(i')$ that of the occurrence of E_i'. For convenience we assume that for any i $p_i(1 - p_i) \neq 0$; events with the exceptional probabilities 0 or 1 may evidently be left out of account.

Let $p(\nu_1 \cdots \nu_k)$ denote the probability of the occurrence of the conjunction $E_{\nu_1} \cdots E_{\nu_k}$ and let $p(\mu_1 \cdots \mu_h, \nu_1 \cdots \nu_k)$ denote the probability of the occurrence of $E_{\nu_1} \cdots E_{\nu_k}$, on the hypothesis that $E_{\mu_1} \cdots E_{\mu_h}$ have occurred. The μ's or ν's may be accented.

Definition 1: *If $p(\nu_1, \nu_2) > p(\nu_2)$, we say that the occurrence of the event E_{ν_1} is favorable to the occurrence of the event E_{ν_2}, or simply that E_{ν_1} is favorable to E_{ν_2}.*

[1] "On the probability of the occurrence of at least m events among n arbitrary events," *Annals of Math. Stat.* Vol. 12 (1941), pp. 328–338.

If $p(\nu_1, \nu_2) = p(\nu_2)$, *we say that* E_{ν_1} *is indifferent to* E_{ν_2}. *If* $p(\nu_1, \nu_2) < p(\nu_2)$, *we say that* E_{ν_1} *is unfavorable to* E_{ν_2}.

Thus the relations "favorableness," "indifference," and "unfavorableness" are mutually exclusive and together exhaustive. We state the following immediate consequences:

(i) Reflexity: An event is favorable to itself; in fact, $p(\nu, \nu) = 1 > p(\nu)$.

(ii) Symmetry: If E_1 is favorable (indifferent, unfavorable) to E_2, then E_2 is favorable (indifferent, unfavorable) to E_1. In fact, we have

$$p(1)p(1, 2) = p(12) = p(2)p(2, 1),$$

$$\frac{p(1, 2)}{p(2)} = \frac{p(2, 1)}{p(1)}.$$

Thus $p(1, 2) \gtreqless p(2)$ is equivalent to $p(2, 1) \gtreqless p(1)$.

In particular, if E_1 is indifferent to E_2, then so is E_2 to E_1. They are then usually said to be independent of each other.

(iii) If E_1 is favorable (indifferent, unfavorable) to E_2, then E_1' is unfavorable (indifferent, favorable) to E_2. For, we have

$$p(1)p(1, 2) + p(1')p(1', 2) = p(12) + p(1'2) = p(2),$$

whence

$$p(1')p(1', 2) = p(2) - p(1)p(1, 2).$$

On the other hand,

$$p(1')p(2) = [1 - p(1)]p(2) = p(2) - p(1)p(2).$$

Since by assumption $p(1')p(2) \neq 0$, we have

$$\frac{p(1', 2)}{p(2)} = \frac{p(2) - p(1)p(1, 2)}{p(2) - p(1)p(2)}.$$

Thus

$$p(1', 2) \lesseqgtr p(2) \text{ according as } p(1, 2) \gtreqless p(2).$$

For the sake of brevity we introduce the following symbolic notation:

$$E_1/E_2 = \begin{cases} 1, \text{ if } E_1 \text{ is favorable to } E_2 \\ 0, \text{ if } E_1 \text{ is indifferent to } E_2 \\ -1, \text{ if } E_1 \text{ is unfavorable to } E_2 . \end{cases}$$

Then by (ii) and (iii) we have

$$E_1/E_2 = E_2/E_1,$$

$$E_1'/E_2 = E_2/E_1' = E_1/E_2' = E_2'/E_1 = -(E_1/E_2),$$

$$E_1'/E_2' = E_2'/E_1' = E_1/E_2,$$

analogous to the rules of signs in the multiplication of integers.

(iv) Non-transitivity: If E_1 is favorable to E_2, and E_2 is favorable to E_3, it does not necessarily follow that E_1 is favorable to E_3; in fact, it may happen that E_1 is unfavorable to E_3. For instance, imagine 11 identical balls in a bag marked respectively with the numbers

$$-11, -10, -3, -2, -1, 2, 4, 6, 11, 13, 16.$$

Let a ball be drawn at random. Let

E_1 = (the event of the number on the ball being positive)

E_2 = (the event of the number on the ball being even)

E_3 = (the event of the number on the ball being of 1 digit)

We have

$$p(1, 2) = \tfrac{4}{6} > \tfrac{6}{11} = p(2),$$

$$p(2, 3) = \tfrac{4}{6} > \tfrac{6}{11} = p(3),$$

$$p(1, 3) = \tfrac{1}{2} < \tfrac{6}{11} = p(3).$$

(v) It may happen that $E_1/E_3 = 1$, $E_2/E_3 = 1$, but $E_1E_2/E_3 = -1$. In the example above,

$$p(2, 1) = \tfrac{4}{6} > \tfrac{6}{11} = p(1),$$

$$p(3', 1) = \tfrac{3}{5} > \tfrac{6}{11} = p(1),$$

$$p(23', 1) = \tfrac{1}{2} < \tfrac{6}{11} = p(1).$$

(vi) It may happen that $E_1/E_2 = 1$, $E_1/E_3 = 1$, but $E_1/E_2E_3 = -1$. Example:

$$p(1, 2) = \tfrac{4}{6} > \tfrac{6}{11} = p(2),$$

$$p(1, 3') = \tfrac{1}{2} > \tfrac{5}{11} = p(3'),$$

$$p(1, 23') = \tfrac{1}{6} < \tfrac{2}{11} = p(23').$$

(vii) It may happen that $E_1/E_3 = 1$, $E_2/E_3 = 1$, but the disjunction $(E_1 + E_2)/E_3 = -1$. For, by (v) we know that there exist events E_1', E_2', E_3' such that

$$E_1'/E_3' = 1, \qquad E_2'/E_3' = 1, \qquad E_1'E_2'/E_3' = -1.$$

Hence by (iii) there exist events E_1, E_2, E_3 such that

$$E_1/E_3 = 1, \qquad E_2/E_3 = 1, \qquad (E_1'E_2')'/E_3 = -1.$$

But $(E_1'E_2')' = E_1 + E_2$. Thus the last relation is $(E_1 + E_2)/E_3 = -1$.

(viii) It may happen that $E_1/E_2 = 1$, $E_1/E_3 = 1$, but $E_1/(E_2 + E_3) = -1$. This follows from (vi) as (vii) follows from (v).

After all these negative results in (iv)–(viii), we see that we cannot expect to go far without making stronger assumptions regarding the nature of inter-

dependence between the events in the set. Firstly, in view of (iv), we shall restrict ourselves to consideration of a set of events in which each event is favorable to every other. Secondly, in view of (v), we shall only consider the case where the "favorableness," as defined above, shall be cumulative in its effect, that is to say, the more events favorable to a given event have been known to occur, the more probable this given event shall be esteemed. We formulate these two conditions in mathematical terms, as follows:

DEFINITION 2: *A set of events $E_1, \cdots, E_n, \cdots$ is said to be strongly mutually favorable (in the first sense) if, for every integer h and every set of distinct indices (positive integers) $\mu_1, \cdots, \mu_h$ and ν we have*

$$p(\mu_1 \cdots \mu_h, \nu) > p(\mu_1 \cdots \mu_{h-1}, \nu).$$

This definition requires that there exist no implication relation between any event and any conjunction of events in the set; in particular, that the events are all distinct. It would be more convenient to consider the relation "favorable or indifferent to." This will be done later on. The present definitions have the advantage of being logically clear cut and also that of yielding unambiguous inequalities.

From Definition 2 we deduce the following consequences:

(1) If the set $(\mu_1^*, \cdots, \mu_j^*)$ is a sub-set of $(\mu_1, \cdots, \mu_h)$, we have

$$p(\mu_1 \cdots \mu_h, \nu) > p(\mu_1^* \cdots \mu_j^*, \nu).$$

(2) For any positive integer k and any two sets $(\nu_1, \cdots, \nu_k)$ and $(\mu_1, \cdots, \mu_h)$ where all the indices are distinct, we have

$$p(\mu_1 \cdots \mu_h, \nu_1 \cdots \nu_k) > p(\mu_1 \cdots \mu_{h-1}, \nu_1 \cdots \nu_k).$$

More generally, we have as in (1),

$$p(\mu_1 \cdots \mu_h, \nu_1 \cdots \nu_k) > p(\mu_1^* \cdots \mu_j^*, \nu_1 \cdots \nu_k).$$

PROOF: We have only to prove the first inequality. For $k = 1$ this is the assumption in Definition 2. Suppose that the inequality holds for $k - 1$, we shall prove that it holds for k, too:

$$\frac{p(\mu_1 \cdots \mu_h, \nu_1 \cdots \nu_k)}{p(\mu_1 \cdots \mu_{h-1}, \nu_1 \cdots \nu_k)} = \frac{p(\mu_1 \cdots \mu_{h-1})p(\mu_1 \cdots \mu_h)p(\mu_1 \cdots \mu_h, \nu_1 \cdots \nu_k)}{p(\mu_1 \cdots \mu_h)p(\mu_1 \cdots \mu_{h-1})p(\mu_1 \cdots \mu_{h-1}, \nu_1 \cdots \nu_k)}$$

$$= \frac{p(\mu_1 \cdots \mu_{h-1})p(\mu_1 \cdots \mu_h \nu_1 \cdots \nu_k)}{p(\mu_1 \cdots \mu_h)p(\mu_1 \cdots \mu_{h-1}\nu_1 \cdots \nu_k)}$$

$$= \frac{p(\mu_1 \cdots \mu_{h-1})p(\mu_1 \cdots \mu_h)p(\mu_1 \cdots \mu_h, \nu_1)p(\mu_1 \cdots \mu_h \nu_1, \nu_2 \cdots \nu_k)}{p(\mu_1 \cdots \mu_h)p(\mu_1 \cdots \mu_{h-1})p(\mu_1 \cdots \mu_{h-1}, \nu_1)p(\mu_1 \cdots \mu_{h-1}\nu_1, \nu_2 \cdots \nu_k)}$$

$$= \frac{p(\mu_1 \cdots \mu_h, \nu_1)}{p(\mu_1 \cdots \mu_{h-1}, \nu_1)} \frac{p(\mu_1 \cdots \mu_h \nu_1, \nu_2 \cdots \nu_k)}{p(\mu_1 \cdots \mu_{h-1}\nu_1, \nu_2 \cdots \nu_k)}$$

$$> \frac{p(\mu_1 \cdots \mu_h \nu_1, \nu_2 \cdots \nu_k)}{p(\mu_1 \cdots \mu_{h-1}\nu_1, \nu_2 \cdots \nu_k)} > 1.$$

Observe that none of the denominators vanish by our original assumption and by Definition 2.

Therefore we see that when the failure in (v) is remedied by our definition, the failure in (vi) is automatically remedied too.

2. THEOREM 1: *Let $n > 1$ and let $E_1, \cdots, E_n, \cdots$ be a set of strongly mutually favorable events (in the first sense). Then we have, for $k = 1, \cdots, n-1$,*

$$\prod_{\nu_1, \cdots, \nu_{k+1}} [p(\nu_1 \cdots \nu_{k+1})]^{\binom{n-1}{k}^{-1}} > \prod_{\nu_1, \cdots, \nu_k} [p(\nu_1 \cdots \nu_k)]^{\binom{n-1}{k-1}^{-1}}$$

where the products extend respectively to all combinations of $k+1$ and k distinct indices out of the indices $1, \cdots, n$.

PROOF. We may assume that the indices are written so that

$$1 \leqq \nu_1 < \nu_2 < \cdots < \nu_{k+1} \leqq n.$$

Taking logarithms, we have

$$\binom{n-1}{k-1} \sum_{\nu_1, \cdots, \nu_{k+1}} \log p(\nu_1 \cdots \nu_{k+1}) > \binom{n-1}{k} \sum_{\nu_1, \cdots, \nu_k} \log p(\nu_1 \cdots \nu_k).$$

Substituting from the obvious formula

$$p(\nu_1 \cdots \nu_k) = p(\nu_1)p(\nu_1, \nu_2)p(\nu_1\nu_2, \nu_3) \cdots p(\nu_1 \cdots \nu_{k-1}, \nu_k),$$

and writing $\log p(\cdots) = q(\cdots)$, the inequality becomes

$$\begin{aligned} &\binom{n-1}{k-1} \Sigma[q(\nu_1) + q(\nu_1, \nu_2) + \cdots + q(\nu_1 \cdots \nu_k, \nu_{k+1})] \\ &\qquad > \binom{n-1}{k} \Sigma[q(\nu_1) + q(\nu_1, \nu_2) + \cdots + q(\nu_1 \cdots \nu_{k-1}, \nu_k)]. \end{aligned} \tag{1}$$

Immediately we observe that the number of terms of the form $q(\nu_1 \cdots \nu_s, \mu)(0 \leqq s \leqq \mu - 1)$ with a fixed μ after the comma in the bracket is the same on both sides, since

$$\binom{n-1}{k-1}\binom{n-1}{k} = \binom{n-1}{k}\binom{n-1}{k-1}. \tag{2}$$

Let the sums of such q's on the left and right of (1) be $\sigma^{(1)} = \sigma^{(1)}(\mu)$ and $\sigma^{(2)} = \sigma^{(2)}(\mu)$ respectively. To prove our theorem it is sufficient to prove that $\sigma^{(1)}(\mu) \geqq \sigma^{(2)}(\mu)$ for every μ and $\sigma^{(1)}(\mu) > \sigma^{(2)}(\mu)$ for at least one μ.

Now the terms in $\sigma^{(1)}$ (or $\sigma^{(2)}$) fall into classes according to the number s of the μ_i's before the comma in the bracket. Let those terms having s μ_i's before the comma belong to the s-th class. It is evident that the number of terms of the s-th class in $\sigma^{(1)}(\mu)$ is equal to

$$\binom{n-1}{k-1}\binom{\mu-1}{s}\binom{n-\mu}{k-s}$$

for $s = 0, 1, \cdots, \mu - 1$; where we make the convention that

$$\binom{0}{0} = 1, \qquad \binom{a}{b} = 0 \quad \text{if } a < b \text{ or if } b < 0.$$

Thus for a fixed μ, when the terms in $\sigma^{(1)}(\mu)$ are classified in the above manner, its total number of terms may be written as the following sum, in which vanishing terms may occur:

$$\binom{n-1}{k-1}\binom{n-1}{k} = \binom{n-1}{k-1}\left\{\binom{\mu-1}{\mu-1}\binom{n-\mu}{k-\mu+1} + \binom{\mu-1}{\mu-2}\binom{n-\mu}{k-\mu+2} + \cdots + \binom{\mu-1}{s}\binom{n-\mu}{k-s} + \cdots + \binom{\mu-1}{0}\binom{n-\mu}{k}\right\}.$$

Similarly the total number of terms in $\sigma^{(2)}(\mu)$ may be written as the following sum:

$$\binom{n-1}{k}\binom{n-1}{k-1} = \binom{n-1}{k}\left\{\binom{\mu-1}{\mu-1}\binom{n-\mu}{k-\mu} + \binom{\mu-1}{\mu-2}\binom{n-\mu}{k-\mu+1} + \cdots + \binom{\mu-1}{s}\binom{n-\mu}{k-s-1} + \cdots + \binom{\mu-1}{0}\binom{n-\mu}{k-1}\right\}.$$

LEMMA 1: *For $0 \leqq s \leqq k$, we have, taking account of our conventions about the binomial coefficients,*

$$(3)\qquad \binom{n-1}{k-1}\binom{n-\mu}{k-s} > \binom{n-1}{k}\binom{n-\mu}{k-s-1} \quad \textit{for} \quad s > (\mu-1)k/n;$$

$$(4)\qquad \binom{n-1}{k-1}\binom{n-\mu}{k-s} \leqq \binom{n-1}{k}\binom{n-\mu}{k-s-1} \quad \textit{for} \quad s \leqq (\mu-1)k/n.$$

PROOF: Suppose $s \geqq k - n + \mu$, then

$$\binom{n-1}{k-1}\binom{n-\mu}{k-s} \gtreqless \binom{n-1}{k}\binom{n-\mu}{k-s-1}$$

according as

$$\frac{k}{n-k} \gtreqless \frac{k-s}{n-\mu-k+s+1},$$

i.e. according as

$$s \gtreqless (\mu-1)k/n.$$

But, since $k < n$ and $\mu \leqq n$, we have

$$n - k - k/n + 1 > (n-k)\mu/n$$
$$(\mu-1)k/n > k - n + \mu - 1$$

so that

$$(\mu - 1)k/n + 1 \geqq k - n + \mu.$$

Therefore if $s > (\mu - 1)k/n$, then $s \geqq (\mu - 1)k/n + 1 \geqq k - n + \mu$, and (3) holds.

Again, if $k - n + \mu \leqq s \leqq (\mu - 1)k/n$, then (4) holds; while if $s < k - n + \mu$, then the left-hand side of (4) vanishes while the right-hand side is non-negative, thus (4) holds for $s \leqq (\mu - 1)k/n$. The lemma is proved.

If we put $(s = 0, 1, \cdots, k)$

$$\binom{n-1}{k-1}\binom{n-\mu}{k-s} - \binom{n-1}{k}\binom{n-\mu}{k-s-1} = d_s,$$

then by Lemma 1,

$$d_s \gtreqless 0 \quad \text{according as} \quad s \gtreqless (\mu - 1)k/n.$$

This means that although the total number of terms of the form $p(\mu_1 \cdots \mu_s, \mu)$ is the same on both sides of (1), the left-hand side is more abundant in terms with larger s while the right-hand side is more abundant in terms with smaller s. Now we have

$$q(\mu_1 \cdots \mu_i, \mu) > q(\mu_1^* \cdots \mu_j^*, \mu)$$

if $i > j$ and if $(\mu_1^* \cdots \mu_j^*)$ is a subset of $(\mu_1 \cdots \mu_i)$. Hence it is natural to suppose that the left-hand side must be greater because it is more abundant in terms of larger values. Unfortunately even if $i > j$, the last inequality is in general not true if the set $(\mu_1^* \cdots \mu_j^*)$ is not a sub-set of $(\mu_1 \cdots \mu_i)$. Therefore we cannot as yet conclude that $\sigma^{(1)} \geqq \sigma^{(2)}$.

To prove that actually we have $\sigma^{(1)} \geqq \sigma^{(2)}$, we make the following "process of compensation":

We have, by (2) and the definition of d_s, the following equality:

$$\binom{\mu-1}{0}d_0 + \binom{\mu-1}{1}d_1 + \cdots + \binom{\mu-1}{\mu-1}d_{\mu-1} = 0.$$

where $d_j = 0$ if $j > k$. Thus

$$d_s \leqq 0 \quad \text{for} \quad s \leqq k(\mu - 1)/n,$$

$$d_s \geqq 0 \quad \text{for} \quad s > k(\mu - 1)/n.$$

Hence

$$\binom{\mu-1}{0}d_0 + \binom{\mu-1}{1}d_1 + \cdots + \binom{\mu-1}{\mu-1}d_{\mu-1} \leqq 0 \tag{5}$$

for $l = 0, 1, \cdots, \mu - 1.$

For the fixed μ, let

$$\rho_l^{(1)} = \binom{n-1}{k-1}\left\{\binom{n-\mu}{k}q_\mu + \binom{n-\mu}{k-1}\sum_{\mu_1<\mu} q(\mu_1, \mu) + \cdots \right.$$
$$\left. + \binom{n-\mu}{k-l}\sum_{\mu_1<\cdots<\mu_l<\mu} q(\mu_1 \cdots \mu_l, \mu)\right\}$$

$$\rho_l^{(2)} = \binom{n-1}{k}\left\{\binom{n-\mu}{k-1}q_\mu + \binom{n-\mu}{k-2}\sum_{\mu_1<\mu} q(\mu_1, \mu) + \cdots \right.$$
$$\left. + \binom{n-\mu}{k-l-1}\sum_{\mu_1<\cdots<\mu_l<\mu} q(\mu_1 \cdots \mu_l, \mu)\right\}$$

so that

$$\rho_{\mu-1}^{(1)} = \sigma^{(1)}(\mu), \qquad \rho_{\mu-1}^{(2)} = \sigma^{(2)}(\mu).$$

For $\mu = 1$, $l = 0$, we have

$$\sigma^{(1)}(1) = \rho_0^{(1)} = \rho_0^{(2)} = \sigma^{(2)}(1).$$

LEMMA 2: *For $\mu > 1$ and $0 \leqq l < \mu - 1$, we have*

$$\sum_{1\leqq\mu_1<\cdots<\mu_l<\mu} q(\mu_1 \cdots \mu_l, \mu) < \frac{l+1}{\mu - l - 1}\sum_{1\leqq\mu_1<\cdots<\mu_{l+1}<\mu} q(\mu_1 \cdots \mu_{l+1}, \mu).$$

PROOF: We have, for any $\nu < \mu$, $\nu \neq \mu_i$ $(i = 1, \cdots, l)$

$$q(\mu_1 \cdots \mu_l\nu, \mu) > q(\mu_1 \cdots \mu_l, \mu).$$

Summing with respect to all such ν's,

$$\sum_\nu q(\mu_1 \cdots \mu_l\nu, \mu) > (\mu - l - 1)q(\mu_1 \cdots \mu_l, \nu).$$

Summing with respect to all $1 \leqq \mu_1 < \cdots < \mu_l < \mu$,

$$\sum_{1\leqq\mu_1<\cdots<\mu_l<\mu}\sum_\nu q(\mu_1 \cdots \mu_l\nu, \mu) = (l+1)\sum_{1\leqq\mu_1<\cdots<\mu_{l+1}<\mu} q(\mu_1 \cdots \mu_{l+1}, \mu)$$
$$> (\mu - l - 1)\sum_{1\leqq\mu_1<\cdots<\mu_l<\mu} q(\mu_1 \cdots \mu_l, \mu).$$

The lemma is proved.

Now we use induction to prove that for $\mu > 1$ and $l = 1, \cdots, \mu - 1$

$$\rho_l^{(1)} - \rho_l^{(2)} > \frac{d_0 + \binom{\mu-1}{1}d_1 + \cdots + \binom{\mu-1}{l}d_l}{\binom{\mu-1}{l}}$$
$$\times \sum_{1\leqq\mu_1<\cdots<\mu_l<\mu} q(\mu_1 \cdots \mu_l, \mu) \geqq 0.$$

This inequality holds for $l = 1$ by Lemma 2. Assume that it holds for l, $(l < \mu - 1)$. Then we have, by (5) and the fact that each $q < 0$,

$$\rho_{l+1}^{(1)} - \rho_{l+1}^{(2)} = \rho_l^{(1)} - \rho_l^{(2)} + d_{l+1} \sum_{1 \leqq \mu_1 < \cdots < \mu_{l+1} < \mu} q(\mu_1 \cdots \mu_{l+1}, \mu)$$

$$> \frac{d_0 + \binom{\mu-1}{1} d_1 + \cdots + \binom{\mu-1}{l} d_l}{\binom{\mu-1}{l}} \sum q(\mu_1 \cdots \mu_l, \mu) + d_{l+1} \sum q(\mu_1 \cdots \mu_{l+1}, \mu)$$

$$\geqq \left(\frac{d_0 + \binom{\mu-1}{1} d_1 + \cdots + \binom{\mu-1}{l} d_l}{\binom{\mu-1}{l}} \frac{l+1}{\mu - l - 1} + d_{l+1} \right) \sum q(\mu_1 \cdots \mu_{l+1}, \mu)$$

$$= \frac{d_0 + \binom{\mu-1}{1} d_1 + \cdots + \binom{\mu-1}{l} d_l + \binom{\mu-1}{l+1} d_{l+1}}{\binom{\mu-1}{l+1}} \sum q(\mu_1 \cdots \mu_{l+1}, \mu) \geqq 0.$$

Therefore, for $\mu > 1$, we have

$$\sigma^{(1)}(\mu) - \sigma^{(2)}(\mu) = \rho_{\mu-1}^{(1)} - \rho_{\mu-1}^{(2)} > 0.$$

Since $n > 1$ and $1 \leqq \mu \leqq n$, there exists a $\mu > 1$. Hence

$$\sum_{\mu=1}^{n} \sigma^{(1)}(\mu) > \sum_{\mu=1}^{n} \sigma^{(2)}(\mu)$$

which is equivalent to the inequality (1).

3. Our next step will be to obtain a generalization of Theorem 1. Consider a derived event defined by a disjunction of a (finite) number of events in the set, as follows:

$$E_{\nu_1} + E_{\nu_2} + \cdots + E_{\nu_m}.$$

We call such a disjunction a disjunction of the m-th order.

DEFINITION 3: *A set of events is said to be strongly mutually favorable in the second sense if for every positive integer m, the derived set of events consisting of all the disjunctions of the m-th order forms a strongly mutually favorable set of events (in the first sense).*

Let $D = D(m)$ denote in general a disjunction of the m-th order; let $p(D_1 \cdots D_h, D)$ denote the probability of the occurrence of the disjunction D, on the hypothesis that the conjunction of the h disjunctions $D_1 \cdots D_h$ has occurred. Then Definition 3 says that for any positive integer h and any set of distinct D's we have

$$p(D_1 \cdots D_h, D) > p(D_1 \cdots D_{h-1}, D).$$

Since a disjunction of the 1st order is an event E, we see that Definition 3 includes Definition 2.

Let $D_m(\nu_1, \cdots, \nu_k)$, $\nu_1 < \cdots < \nu_k$ denote the derived event

$$\prod_{\mu_1, \cdots, \mu_m} (E_{\mu_1} + \cdots + E_{\mu_m})$$

where the product (conjunction) extends to all combinations of m indices out of the indices $\nu_1, \cdots, \nu_k$. Let $p_m^*(\nu_1, \cdots, \nu_k)$ denote the probability of the occurrence of $D_m(\nu_1, \cdots, \nu_k)$. It is seen that $p_1^*(\nu_1, \cdots, \nu_k) = p(\nu_1 \cdots \nu_k)$ in our previous notation.

We merely state Theorem 2, whose proof is analogous to that of Theorem 1 but requires more cumbersome expressions.

THEOREM 2: *Let $n > k \geqq m \geqq 1$ and let $E_1, \cdots, E_n$ be a set of mutually strongly favorable events in the second sense. Then we have*

$$\prod_{1 \leqq \nu_1 < \cdots < \nu_{k+1} \leqq n} [p_m^*(\nu_1, \cdots, \nu_{k+1})]^{\binom{n-m}{k-m+1}^{-1}} > \prod_{1 \leqq \nu_1 < \cdots < \nu_k \leqq n} [p_m^*(\nu_1, \cdots, \nu_k)]^{\binom{n-m}{k-m}^{-1}}.$$

To give an interpretation of $p_m^*(\nu_1, \cdots, \nu_k)$, we prove the symbolic equation between events:

$$D_m = \prod_{\nu_1 \leqq \mu_1 < \cdots < \mu_m \leqq \nu_k} (E_{\mu_1} + \cdots + E_{\mu_m}) = \sum_{\nu_1 \leqq \mu_1 < \cdots < \mu_{k-m+1} \leqq \nu_k} (E_{\mu_1} \cdots E_{\mu_{k-m+1}}) = C_{k-m+1},$$

where product means conjunction and sum means disjunction.

To prove this, we write for a general event E, $E = 1$ when E occurs, $E = 0$ when E does not occur. Now if $C_{k-m+1} = 0$, then at most $k - m$ events among the k given events occur, so that there exist m events such that $E_{\lambda_1} = 0$, $E_{\lambda_2} = 0$, $E_{\lambda_m} = 0$, thus

$$E_{\lambda_1} + E_{\lambda_2} + \cdots + E_{\lambda_m} = 0.$$

Now the last disjunction is contained in D_m as a factor, therefore $D_m = 0$.

Conversely, if $D_m = 0$, at least one of its factors $= 0$, so that there exist m events, such that $E_{\lambda_1} = 0, E_{\lambda_2} = 0, \cdots, E_{\lambda_m} = 0$. Thus at most $k - m$ events out of the k given events occur and so by definition $C_{k-m+1} = 0$. Q.e.d.

From the above it immediately follows that

$$p_m^*(\nu_1, \cdots, \nu_k) = p_{k-m+1}(\nu_1, \cdots, \nu_k)$$

where $p_{k-m+1}(\nu_1, \cdots, \nu_k)$ is defined in the introduction. Then Theorem 2 may be written as

$$\Pi[p_{k-m+2}(\nu_1, \cdots, \nu_{k+1})]^{\binom{n-m}{k-m+1}^{-1}} > \Pi[p_{k-m+1}(\nu_1, \cdots, \nu_k)]^{\binom{n-m}{k-m}^{-1}}$$

or again as

$$\Pi[w_{m-1}(\nu_1', \cdots, \nu_{k+1}')]^{\binom{n-m}{k-m+1}^{-1}} > \Pi[w_{m-1}(\nu_1', \cdots, \nu_k')]^{\binom{n-m}{k-m}^{-1}}$$

where $w_{m-1}(\nu_1', \cdots, \nu_k')$ denotes the probability of the occurrence of at most $m - 1$ events out of the k events $E'_{\nu_1}, \cdots, E'_{\nu_k}$.

REMARK. If in our Definitions 2 and 3 we replace the sign "$>$" by the sign "$\geqq$", then we obtain the inequalities in Theorems 1 and 2 with the sign "$>$" replaced by "$\geqq$". The corresponding set of events thus newly defined will be said to be strongly mutually favorable or indifferent (in the first or second sense).

After this modification, we can include events with the probability 1 in our considerations. Also, the events need no longer be distinct and there may now exist implication relations between events or their conjunctions. This modification is useful for the following application.

4. Consider the divisibility of a random positive integer by the set of positive integers. To each positive integer there corresponds an event, namely the event that the random positive integer is divisible by it. The enumerable set of events

$$E_1, E_2, E_3, E_4, \cdots, E_n, \cdots$$

where E_n = the event of divisibility by n, with the probabilities

$$1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \cdots, \frac{1}{n}, \cdots$$

evidently forms a set of strongly mutually favorable or indifferent events in the second sense.

Again, the enumerable set of events

$$E_2', E_3', E_4', \cdots, E_n', \cdots$$

where E_n' = the event of non-divisibility by n, with the probabilities

$$\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \cdots, \frac{n-1}{n}, \cdots$$

evidently also forms a set of strongly mutually favorable or indifferent events in the second sense.

Hence our Theorem 2 can be applied to both sets and in this way we obtain results which belong properly to the elementary theory of numbers.

We shall content ourselves with indicating a few examples.

Let $\{a_1, \cdots, a_n\}$ denote the least common multiple of the natural numbers $a_1, \cdots, a_n$. Then Theorem 1, when applied to the two sets above, gives respectively

THEOREM 1.1: *Let $a_1, \cdots, a_n$ be any positive integers, then we have, for $k = 1, \cdots, n - 1$*

$$\left(\prod_{1 \leqq \nu_1 < \cdots < \nu_{k+1} \leqq n} \frac{1}{\{a_{\nu_1}, \cdots, a_{\nu_{k+1}}\}}\right)^{\binom{n-1}{k}^{-1}} \geqq \left(\prod_{1 \leqq \nu_1 < \cdots < \nu_k \leqq n} \frac{1}{\{a_{\nu_1}, \cdots, a_{\nu_k}\}}\right)^{\binom{n-1}{k-1}^{-1}}.$$

THEOREM 1.2: *Also we have,*

$$\prod_{1 \leqq \nu_1 < \cdots < \nu_{k+1} \leqq n} \left(1 - \sum_{\nu_1 \leqq \mu_1 \leqq \nu_{k+1}} \frac{1}{a_{\mu_1}} + \sum_{\nu_1 \leqq \mu_1 < \mu_2 \leqq \nu_{k+1}} \frac{1}{\{a_{\mu_1}, a_{\mu_2}\}} - + \cdots + (-1)^{k+1} \frac{1}{\{a_{\nu_1}, \cdots, a_{\nu_{k+1}}\}}\right)^{\binom{n-1}{k}^{-1}}$$

$$\geqq \prod_{1 \leqq \nu_1 < \cdots < \nu_k \leqq n} \left(1 - \sum_{\nu_1 \leqq \mu_1 \leqq \nu_k} \frac{1}{a_{\mu_1}} + \sum_{\nu_1 \leqq \mu_1 < \mu_2 \leqq \nu_k} \frac{1}{\{a_{\mu_1}, a_{\mu_2}\}} - + \cdots + (-1)^k \frac{1}{\{a_{\nu_1}, \cdots, a_{\nu_k}\}}\right)^{\binom{n-1}{k-1}^{-1}}.$$

A trivial corollary of Theorem 1 is

$$p(12 \cdots n) \geqq p_1 p_2 \cdots p_n .$$

Correspondingly we have

$$1 - \sum_{1 \leqq \mu_1 \leqq n} \frac{1}{a_{\mu_1}} + \sum_{1 \leqq \mu_1 < \mu_2 \leqq n} \frac{1}{\{a_{\mu_1}, a_{\mu_2}\}} - + \cdots + (-1)^n \frac{1}{\{a_1, \cdots, a_n\}} \geqq \left(1 - \frac{1}{a_1}\right)\left(1 - \frac{1}{a_2}\right) \cdots \left(1 - \frac{1}{a_n}\right).$$

If we multiply by $a_1 a_2 \cdots a_n$, we get

$$A(a_1, a_2, \cdots, a_n) \geqq (a_1 - 1)(a_2 - 1) \cdots (a_n - 1),$$

where $A(a_1, \cdots, a_n)$ denotes the number of positive integers $\leqq a_1 a_2 \cdots a_n$ that are not divisible by any of the a_i $(i = 1, \cdots, n)$.

This last result, which is almost obvious here, was proved by H. Rohrbach and H. Heilbronn independently.[2] See also my generalization[3] (also obvious from the present point of view) of this result to higher dimensional sets of positive integers and to sets of ideals in any algebraic number field.

---

[2] "Beweis einer zahlentheoretische Ungleichung," *Jour. für Math.*, Vol. 177 (1937), pp. 193–196. "On an inequality in the elementary theory of numbers," *Proc. Camb. Phil. Soc.*, Vol. 33, (1937), pp. 207–209.

[3] "A generalization of an inequality in the elementary theory of numbers," *Jour. für Math.*, Vol. 183 (1941), p. 103.

OBSERVATIONS ON ANALYSIS OF VARIANCE THEORY

By Hilda Geiringer[1]

Bryn Mawr College

One of the important problems of theoretical statistics is the following. Let $x_1, x_2, \cdots x_N$ be the results of N observations; by means of these results we want to test the hypothesis that $V_i(x)$ is the distribution of the ith chance variable x_i. In that situation we often decide to choose a test function $F(x_1, x_2, \cdots x_N)$ and to determine the distribution of F under the above assumption. By means of this distribution we compute the probability of $\xi_1 \leqq F \leqq \xi_2$ and compare this result with the observed value of F.

Suppose there are m groups, each of n observations on $m \cdot n$ chance variables $x_{\mu\nu}$. We may test hypotheses regarding the mn distributions of the $x_{\mu\nu}$ in the way just mentioned. In analysis of variance theory we often use as test functions certain quadratic forms s_w^2 and s_a^2 ("variance within" and "among classes") and their quotient (multiplied by $m(n-1)/(m-1)$), usually denoted by z. Its distribution has been investigated by R. A. Fisher [2] under the assumption that the chance variables are mutually independent and subject to the same normal law. "The five per cent and one per cent points of this distribution have been tabulated by R. A. Fisher and are used to test, whether these two estimates of the same magnitude are significantly different. One gets thus a test of significance *to test whether our sample is a random sample from a homogeneous normal population or not.*[2] If the probability of a certain z-value is too small we shall reject the hypothesis that the sample is a random sample from a homogeneous normal population" [5].

The use of Fisher's z-test is also recommended if we may reasonably assume that the theoretical distributions are approximately normal. "Unless some rather startling lack of normality is known or suspected analysis of variance may be used with confidence." This last remark can be understood by considering that, as we will see in detail, some of the basic results of our theory are independent of the normality of the populations. It is however this assumption of normality which makes possible the complete and elegant solution of the problem of distribution obtained by R. A. Fisher.

If it is *not* possible to determine the exact distribution of a test function under sufficiently general assumptions we may:

a) make simple and particular assumptions concerning the populations

b) investigate an asymptotic solution of the problem, i.e. determine the distributions of the test functions for large samples,[3] or

c) study the mathematical expectations and the variances of the test functions

[1] Research under a grant in aid of the American Philosophical Society.

[2] My italics.

[3] cf. statement (a) page 355.

for small samples under appropriately general assumptions regarding the populations (this should be done independently of concepts of estimation, unbiased estimate etc.).

This last procedure provides us with tests which suffice in actual practice.[4]

It is well known that the expectations of the two forms $s_a^2/(m-1)$ and $s_w^2/m(n-1)$ are the same even if the populations are not normal, but equal each other (*Bernoulli* series). In addition we shall prove the theorem, familiar in case of the Lexis quotient [9], that under these conditions the *expectation of their quotients equals unity* (section 1, (b)). The next step consists in investigating certain inequalities characteristic of *Lexis* or *Poisson* series. The different criteria will be completed by the computation of the respective variances (Section 1, (c)).

In addition to the above mentioned test functions other symmetrical test functions have been considered [5]. In studying these we shall again assume general populations. It will be seen that the Lexis as well as the Poisson series may be characterized by equalities (instead of inequalities) (Section 2, (a)), and we can generalize our theorem on the expectation of the quotient (Section 2, (b)) to this case. Then the variances of these test functions will be investigated.

It seems worthwhile *to omit the assumption of independence* of the chance variables and to study different kinds of mutual dependence. These investigations lead to interesting relations among the expectations[5] (Section 2, (c)). They seem to be related to Fisher's "intraclass correlation" and to supplement his idea.

Most of the results of Sections 1 and 2 can be generalized to the analysis of covariance (section 3).

1. Variance within and among classes.

(*a*). *The test functions.* Let $x_{\mu\nu}$ ($\mu = 1, \cdots m$; $\nu = 1, \cdots n$) be $m \cdot n$ chance variables and put

$$\begin{aligned} a_\mu &= \frac{1}{n}\sum_{\nu=1}^{n} x_{\mu\nu}, \qquad \bar{a}_\nu = \frac{1}{m}\sum_{\mu=1}^{m} x_{\mu\nu}, \\ a &= \frac{1}{mn}\sum_{\mu=1}^{m}\sum_{\nu=1}^{n} x_{\mu\nu} = \frac{1}{m}\sum_{\mu=1}^{m} a_\mu = \frac{1}{n}\sum_{\nu=1}^{n} \bar{a}_\nu . \end{aligned} \tag{1}$$

[4] The important paper of Irwin [5] assumes normality of the populations. H. L. Rietz [8] computes the expectations of s_a^2 and s_w^2 under rather general assumptions for the populations and considers the cases of Bernoulli, Lexis, Poisson series, but does not consider tests of significance; nor does he consider the symmetric test functions (section 2 of this paper). In later papers on our subject the assumption of normal and independent populations recurs. Another approach [11] in the problem of analysis of variance is to use ranks instead of the actual values (this has been pointed out by the referee to the author, who is very grateful for this comment).

[5] They generalize previous results of the author.

We then introduce the three quadratic forms

$$(2)\quad s^2 = \sum_\mu \sum_\nu (x_{\mu\nu} - a)^2; \qquad s_a^2 = n \sum_\mu (a_\mu - a)^2; \qquad s_w^2 = \sum_\mu \sum_\nu (x_{\mu\nu} - a_\mu)^2,$$

with the respective ranks (degrees of freedom)

$$(3)\qquad r = mn - 1, \qquad r_a = m - 1, \qquad r_w = m(n - 1).$$

Then we have

$$(4)\qquad s^2 = s_a^2 + s_w^2, \qquad r = r_a + r_w.$$

The $m \cdot n$ theoretical distributions are assumed in this section to be independent of each other. Let $V_{\mu\nu}(x)$ be the probability that $x_{\mu\nu} \leqq x$ and

$$(5)\qquad \alpha_{\mu\nu} = \int x\, dV_{\mu\nu}(x), \qquad \sigma_{\mu\nu}^2 = \int (x - \alpha_{\mu\nu})^2\, dV_{\mu\nu}(x),$$

where the integrals are *Stieltjes* integrals; thus the $V_{\mu\nu}(x)$ may be e.g. general arithmetical or geometrical distributions.[6]

Let us compute the mathematical expectation of the three test functions with respect to the $m \cdot n$-dimensional distribution:

$$V_{11}(x_{11}) V_{12}(x_{12}) \cdots V_{mn}(x_{mn}).$$

$$(6)\quad E[F(x_{11}, \cdots x_{mn})] = \int \cdots \int F(x_{11}, \cdots x_{mn})\, dV_{11}(x_{11}) \cdots dV_{mn}(x_{mn}).$$

We have then

$$(7)\qquad E\left[\frac{s^2}{mn - 1}\right] = \frac{1}{mn} \Sigma\Sigma\sigma_{\mu\nu}^2 + \frac{1}{mn - 1} \Sigma\Sigma(\alpha_{\mu\nu} - \alpha)^2,$$

$$(8)\qquad E\left[\frac{s_{a^2}}{m - 1}\right] = \frac{1}{mn} \Sigma\Sigma\sigma_{\mu\nu}^2 + \frac{1}{m - 1} \cdot n\Sigma(\alpha_\mu - \alpha)^2,$$

$$(9)\qquad E\left[\frac{s_w^2}{m(n - 1)}\right] = \frac{1}{mn} \Sigma\Sigma\sigma_{\mu\nu}^2 + \frac{1}{m(n - 1)} \Sigma\Sigma(\alpha_{\mu\nu} - \alpha_\mu)^2.$$

From these equalities we deduce:

1. *If the $m \cdot n$ theoretical mean values $\alpha_{\mu\nu}$ are all equal* (Bernoulli series), *then the expectations in* (6), (7), (8) *are equal; i.e.*

$$(10)\qquad E_B\left(\frac{s^2}{mn - 1}\right) = E_B\left(\frac{s_a^2}{m - 1}\right) = E_B\left(\frac{s_w^2}{m(n - 1)}\right).$$

2. *If the $\alpha_{\mu\nu}$ are equal "by rows" but differ from row to row* (Lexis series), i.e. $\alpha_{\mu\nu} = \alpha_\mu$ *but* $\alpha_\mu \neq \alpha$. *Then*

---

[6] $V_{\mu\nu}(x)$ is a monotone non-decreasing function. Hence it has at most a denumerable set of ordinary jump discontinuities; at such a point it is continuous to the right but not to the left. Moreover it possesses a finite derivative $v_{\mu\nu}(x)$ almost everywhere.

$$E_L\left[\frac{s_a^2}{m-1} - \frac{s^2}{mn-1}\right] = \frac{mn(n-1)}{(m-1)(mn-1)} \sum_\mu (\alpha_\mu - \alpha)^2 > 0, \tag{11}$$

$$E_L\left[\frac{s_a^2}{m-1} - \frac{s_w^2}{m(n-1)}\right] = \frac{n}{m-1} \sum_\mu (\alpha_\mu - \alpha)^2 > 0. \tag{12}$$

3. *If the $\alpha_{\mu\nu}$ are equal "by columns" but differ from column to column* (Poisson series), *then $\alpha_{\mu\nu} = \bar{\alpha}_\nu$; $\alpha_\mu = \alpha$ and*

$$E_P\left[\frac{s_a^2}{m-1} - \frac{s^2}{mn-1}\right] = -\frac{m}{mn-1} \sum_\nu (\bar{\alpha}_\nu - \alpha)^2 < 0, \tag{13}$$

$$E_P\left[\frac{s_a^2}{m-1} - \frac{s_w^2}{m(n-1)}\right] = -\frac{1}{n-1} \sum_\nu (\bar{\alpha}_\nu - \alpha)^2 < 0. \tag{14}$$

In the Lexis theory[7] we speak of *normal*, *supernormal* or *subnormal* dispersion depending on whether the *observed* value of $\dfrac{s_a^2}{m-1}$ is *equal*, *greater* or *less* than that of $\dfrac{s^2}{mn-1}$ and we usually consider the quotient

$$L = \frac{s_a^2}{m-1} \Big/ \frac{s^2}{mn-1}. \tag{15}$$

In analysis of variance theory we usually compare $s_a^2/(m-1)$ (variance among rows) with $s_w^2/m(n-1)$ (variance within rows) and introduce the quotient

$$V = \frac{s_a^2}{m-1} \Big/ \frac{s_w^2}{m(n-1)}. \tag{16}$$

It follows from (4): *If $L \gtreqless 1$ then $V \gtreqless 1$ and conversely.* We may therefore speak of *normal* or *non-normal dispersion* with respect either to L or to V.

The results given by equations (10)–(14) can be expressed as follows: *If the $m \cdot n$ theoretical distributions are all equal the mathematical expectation of s^2/r, of s_a^2/r_a and of s_w^2/r_w are equal. In the case of a Lexis series the expectation of s_a^2/r_a is greater than s^2/r and greater than s_w^2/r_w and in the case of a Poisson series the opposite is true.*

We generally use these facts in order to *make inferences about the unknown populations from the observed values of our test functions $V_{\mu\nu}(x)$.* If e.g., the observed value of s_a^2/r_a is "significantly"[8] greater than that of s^2/r we may assume that the theoretical distributions form a Lexis series. But of course such a significant deviation can also be explained by quite different assumptions regarding the populations (see Section 2, (c)).

(*b*). *Mathematical expectation of the quotient of the test functions.* We are going to prove in this section a theorem of some mathematical interest. This theorem is a generalization of an analogous theorem in the Lexis theory [9].

[7] The relation between these considerations and the Lexis theory will be dealt with in another paper.

[8] The meaning of the word "significantly" has of course still to be explained.

We have seen (10) that the mathematical expectations, defined by (6), of the three test functions

$$S = \frac{s^2}{mn - 1}, \qquad S' = \frac{s_a^2}{m - 1}, \qquad S'' = \frac{s_w^2}{m(n - 1)},$$

are equal *if the $m \cdot n$ populations are equal (i.e. have identical distributions). We will show that even in this case*

$$E\left(\frac{S'}{S}\right) = 1, \qquad E\left(\frac{S''}{S}\right) = 1. \tag{17}$$

Let us put $m \cdot n = N$, and let the N chance variables be arranged in a one-dimensional sequence. As S' and S are of second degree in the x_ν ($\nu = 1, 2, \cdots N$) we may write

$$S' - S = A + \sum B_\nu x_\nu + \sum C_\nu x_\nu^2 + \sum_{\nu \neq \rho} D_{\nu\rho} x_\nu x_\rho$$

where the A, B_ν, C_ν and $D_{\nu\rho}$ are constants. Now form the expectation, defined by (6), of $(S' - S)$ under the assumption that the N populations are equal $V_\nu(x) = V(x)$ ($\nu = 1 \cdots N$). Denoting by α and σ^2 the mean value and variance of $V(x)$ and putting $\Sigma B_\nu = B$, $\Sigma C_\nu = C$, $\Sigma D_{\nu\rho} = D$ we find

$$E(S' - S) = A + B\alpha + C(\sigma^2 + \alpha^2) + D\alpha^2 = 0.$$

And as this equality holds for an arbitrary distribution $V(x)$, we deduce that $A = B = C = D = 0$. Let us then compute under the same assumption the expectation of $(S' - S)/S$. Now the expectations of $1/S$, x_ν/S, x_ν^2/S, $x_\nu x_\rho/S$, take the place of the expectations of 1, x_ν, x_ν^2, $x_\nu x_\rho$. But these new expectations are also *independent of the index* ν, because of the *equality of the N populations* and the *symmetry* of S in the N variables $x_1, \cdots x_N$. Hence we may write

$$E\left(\frac{1}{S}\right) = \mu_0, \qquad E\left(\frac{x_\nu}{S}\right) = \mu_1, \qquad E\left(\frac{x_\nu^2}{S}\right) = \mu_2, \qquad E\left(\frac{x_\nu x_\rho}{S}\right) = \mu_3,$$

and we find

$$E\left(\frac{S' - S}{S}\right) = E\left(\frac{S'}{S} - 1\right) = A\mu_0 + B\mu_1 + C\mu_2 + D\mu_3 = 0,$$

because $A = B = C = D = 0$. Hence $E(S'/S) = 1$.

We may prove in the same way that $E(S''/S) = 1$.

We have however proved (17) only under the assumption that all the N *populations are equal*, whereas (10) is true under the mere hypothesis that the mean values of the populations $V_\nu(x)$ are the same.

(*c*). *The variances of the test functions.* The distribution of our test functions and of their quotients V or L have been determined and tabulated by R. A. Fisher under the hypothesis that the $m \cdot n$ chance variables are *independent* and obey the same normal Gaussian law. Consequently by means of Fisher's distri-

bution we can test only the hypothesis that the theoretical populations have *both these* properties.

If in a statistical problem it is not possible to determine the exact distributions of the test functions under sufficiently general assumptions regarding the populations, one of the following procedures is frequently used:
a) one tries to find an *asymptotic* solution of the problem, i.e. to determine the distribution of the test functions in question for *large samples.* The distribution of the analysis of variance quotient, as n tends to infinity, has been established by W. G. Madow [6]. The same problem for the Lexis quotient was solved as early as 1873 by Helmert [4]. *As m tends to infinity* the limiting distribution is a Gaussian distribution, which follows from general theorems of v. Mises [7].
b) For *small* samples, i.e. if m and n are finite we may determine the expectations and the variances of the test functions for appropriately general populations and establish in this way a test of significance.

In this section we shall compute the variances of our test functions. Let us first assume arbitrary but equal populations $V_\nu(x) = V(x)$ and denote by M_i the ith moment about the mean of $V(x)$:

$$\mathrm{M}_i = \int (x - \alpha)^i \, dV(x), \qquad (i = 1, 2, \cdots),$$
$$(18) \qquad \alpha = \int x \, dV(x), \qquad \mathrm{M}_2 = \sigma^2.$$

Then we find immediately the variance of $S = \dfrac{s^2}{mn - 1}$ using a well-known formula for the variance of a sample variance

$$(19) \qquad \mathrm{Var}\left\{\frac{s^2}{mn-1}\right\} = \mathrm{Var}\left\{\frac{\Sigma\Sigma(x_{\mu\nu} - a)^2}{mn-1}\right\} = \frac{1}{mn}\left\{\mathrm{M}_4 - \frac{mn-3}{mn-1}\mathrm{M}_2^2\right\}.$$

If we need the analogous variance in case of different populations we let

$$t^2 = \sum_{\rho=1}^{r} (y_\rho - b)^2 \qquad \text{where } b = \frac{1}{r}(y_1 + \cdots + y_r)$$

and let $V_\rho(y)$, $(\rho = 1, \cdots r)$, be r populations, and

$$\beta_\rho = \int y \, dV_\rho(y), \qquad \frac{1}{r}\sum_{\rho=1}^{r} \beta_\rho = \beta,$$
$$\int (y - \beta_\rho)^i \, dV_\rho(y) = \mu_i^{(\rho)}, \qquad (i = 1, 2, \cdots, \rho = 1, 2, \cdots r), \mu_2^{(\rho)} = \sigma_\rho^2.$$

Then the following formula may be used:

$$(20) \qquad \mathrm{Var}\,(t^2) = \left(\frac{r-1}{r}\right)^2 \sum_{\rho=1}^{r} [\mu_4^{(\rho)} - \sigma_\rho^4]$$
$$+ 4\,\frac{r-1}{r}\sum_{\rho=1}^{r} \mu_3^{(\rho)}(\beta_\rho - \beta)^2 + 4\sum_{\rho=1}^{r} \sigma_\rho^2(\beta_\rho - \beta)^2 + \frac{4}{r^2}\sum_{\rho<\tau} \sigma_\rho^2 \sigma_\tau^2.$$

We may check (20) by putting the $V_\rho(y)$ all equal to $V(y)$ and find

$$\text{Var}\,(t^2) = \frac{r-1}{r}\,[(r-1)\mu_4 - (r-3)\sigma^4], \tag{20'}$$

in accordance with (19).

In order to determine the variance of s_a^2 by means of these formulae we consider $\frac{1}{m}\sum_\mu (a_\mu - a)^2$ as a sample variance. The n distributions in the nth row are $V_{\mu 1}(x)$, $V_{\mu 2}(x)$, $\cdots$ $V_{\mu n}(x)$. Or, if we assume that they are all equal, simply $V(x) = V(x_{\mu\nu})$. Let us put $\frac{1}{n}\,x_{\mu\nu} = z_{\mu\nu}$ and $V(x_{\mu\nu}) = V'(z_{\mu\nu})$, and denote by $W(a_\mu)$ the distribution of the average of the elements in the μth row:

$$W(a_\mu) = \int \cdots \int dV'(z_{\mu 1})\,dV'(z_{\mu 2}) \cdots dV'(z_{\mu,n-1})V'(a_\mu - z_{\mu 1} - \cdots - z_{\mu,n-1}).$$

There is such a distribution for each row, and we have to find the variance of $\sum_\mu (a_\mu - a)^2$ with respect to the combination of these m distributions. In order to be able to apply (20′) we need the second and fourth moments of these distributions. We have for the mean value α' of $W(a_\mu)$:

$$\alpha' = n\cdot(\text{mean value of } V') = n\cdot\frac{1}{n}\,\alpha_\mu = \alpha$$

and for the variance μ_2' of $W(a_\mu)$: $\mu_2' = \frac{\sigma^2}{n}$. We still need μ_4'. By repeated use of the formula

$$\begin{aligned}\iint [(x_1 - a_1) + (x_2 - a_2)]^4\,dV(x_1)\,dV(x_2)&\\ = \int (x_1 - a_1)^4\,dV(x_1) &+ \int (x_2 - a_2)^4\,dV(x_2)\\ &+ 6\int (x_1 - a_1)^2\,dV(x_1)\int (x_2 - a_2)^2\,dV(x_2),\end{aligned}$$

and of the fact that $W(a_\mu)$ is simply the distribution of the sum of n variables $z_{\mu\nu}$ we get:

$$\mu_4' = \frac{1}{n^4}\left(n\mathrm{M}_4 + 6\,\frac{n(n-1)}{2}\,\mathrm{M}_2^2\right) = \frac{1}{n^3}\,(\mathrm{M}_4 + 3(n-1)\mathrm{M}_2^2)$$

where M_4 and M_2 are the values introduced in (18).

We now apply (20′) and get

$$\text{Var}\,[\Sigma(a_\mu - a)^2] = \frac{m-1}{m}\,[(m-1)\mu_4' - (m-3)\mu_2'].$$

and substituting the values of μ_2' and μ_4', we find by an easy computation the final result:

$$(21) \qquad \operatorname{Var}\left\{\frac{n}{m-1}\,\Sigma(a_\mu - a)^2\right\} = \frac{1}{mn}\,(\mathrm{M}_4 - 3\mathrm{M}_2^2) + \frac{2}{m-1}\,\mathrm{M}_2^2.$$

If we compare this last formula with (19) we see that the right side in (21) is of order $1/m$, whereas that in (19) is of order $1/mn$. Therefore, for sufficiently large values of n, s^2/r will be "more exact" than s_a^2/r_a. In some presentations of the Lexis theory it is implied that the value s_a^2/r_a is to be compared with the theoretical or exact value s^2/r; we may see a certain justification for this idea in the result just mentioned. This may lead us also to use s^2/r as an unbiased estimate of the unknown population variance if $\alpha_{\mu\nu} = \alpha$ (see (7) and (8)).

By means of the simple formulae (19) and (21) we can now easily test whether the values of s^2/r and s_a^2/r_a whose expectations are equal in case of equal populations differ significantly from each other. Of course we must compute as usual approximate values of M_2 and M_4 from the observations. If n is comparatively large—as it usually is e.g. in the Lexis theory—only the term $\dfrac{2}{m-1}\,\mathrm{M}_2^2$ will be significant. If the hypothetical population is Gaussian ($\mathrm{M}_4 = 3\mathrm{M}_2^2$) the right side of (21) reduces to $\dfrac{2}{m-1}\,\mathrm{M}_2^2$ and that of (19) to $\dfrac{2\mathrm{M}_2^2}{mn-1}$; hence these variances are in the ratio of $\dfrac{1}{r_a}\Big/\dfrac{1}{r}$, as one might expect.

2. Symmetric Test Functions.

(*a*). *New equalities for Lexis and Poisson series.* In Section 1, starting with the formula $s^2 = s_a^2 + s_w^2$ we used the test functions s^2/r, s_a^2/r_a, s_w^2/r_w. This implied a difference between rows and columns, which is often justified, e.g. in the Lexis theory. The following decomposition of s^2 is symmetric with respect to rows and columns. Let

$$(1) \qquad \begin{gathered} \frac{1}{n}\sum_{\nu=1}^{n} x_{\mu\nu} = a_\mu, \qquad \frac{1}{m}\sum_{\mu=1}^{m} x_{\mu\nu} = \bar{a}_\nu, \\ \frac{1}{mn}\sum_{\mu=1}^{m}\sum_{\nu=1}^{n} x_{\mu\nu} = \frac{1}{m}\sum_{\mu=1}^{m} a_\mu = \frac{1}{n}\sum_{\nu=1}^{n} \bar{a}_\nu = a, \end{gathered}$$

and

$$(2) \qquad \begin{aligned} s^2 &= \Sigma\Sigma(x_{\mu\nu} - a)^2, & s_a^2 &= n\Sigma(a_\mu - a)^2, & s_w^2 &= \Sigma\Sigma(x_{\mu\nu} - a_\mu)^2 \\ S^2 &= \Sigma\Sigma(x_{\mu\nu} - a_\mu - \bar{a}_\nu + a)^2, & \bar{s}_a^2 &= m\Sigma(\bar{a}_\nu - a)^2, & \bar{s}_w^2 &= \Sigma\Sigma(x_{\mu\nu} - \bar{a}_\nu)^2 \end{aligned}$$

with the respective ranks

$$(3) \qquad \begin{aligned} r &= mn - 1, & r_a &= m - 1, & r_w &= m(n-1), \\ R &= (m-1)(n-1), & \bar{r}_a &= n - 1, & \bar{r}_w &= n(m-1). \end{aligned}$$

Then

$$(5) \qquad s^2 = s_a^2 + \bar{s}_a^2 + S^2 = s_a^2 + s_w^2 = \bar{s}_a^2 + \bar{s}_w^2$$

and

$$r = r_a + \bar{r}_a + R = r_a + r_w = \bar{r}_a + \bar{r}_w\,. \tag{6}$$

We find the expectations of these forms under the assumptions, of arbitrary populations $V_{\mu\nu}(x)$ which are independent and different from each other. We then specialize for Bernouilli series, Lexis and Poisson series of populations respectively. Denoting by $\alpha_{\mu\nu}$ and $\sigma^2_{\mu\nu}$ the mean value and variance of $V_{\mu\nu}(x)$ and by

$$\alpha_\mu = \frac{1}{n}\sum_\nu \alpha_{\mu\nu}, \qquad \bar{\alpha}_\nu = \frac{1}{m}\sum_\mu \alpha_{\mu\nu}, \qquad \alpha = \frac{1}{m}\sum \alpha_\mu = \frac{1}{n}\sum \bar{\alpha}_\nu, \tag{6}$$

we find for the expected values defined in (6) Section 1:

$$\begin{aligned}
E\left[\frac{s^2}{mn-1}\right] &= \frac{1}{mn}\,\Sigma\Sigma\sigma^2_{\mu\nu} + \frac{1}{mn-1}\,\Sigma\Sigma(\alpha_{\mu\nu} - \alpha)^2,\\
E\left[\frac{s_a^2}{m-1}\right] &= \frac{1}{mn}\,\Sigma\Sigma\sigma^2_{\mu\nu} + \frac{1}{m-1}\,n\Sigma(\alpha_\mu - \alpha)^2,\\
E\left[\frac{\bar{s}_a^2}{n-1}\right] &= \frac{1}{mn}\,\Sigma\Sigma\sigma^2_{\mu\nu} + \frac{1}{n-1}\,m\Sigma(\bar{\alpha}_\nu - \alpha)^2,\\
E\left[\frac{S^2}{(m-1)(n-1)}\right] &= \frac{1}{mn}\,\Sigma\Sigma\sigma^2_{\mu\nu} + \frac{1}{(m-1)(n-1)}\,\Sigma\Sigma(\alpha_{\mu\nu} - \alpha_\mu - \bar{\alpha}_\nu + \alpha)^2,\\
E\left[\frac{s_w^2}{m(n-1)}\right] &= \frac{1}{mn}\,\Sigma\Sigma\sigma^2_{\mu\nu} + \frac{1}{m(n-1)}\,\Sigma\Sigma(\alpha_{\mu\nu} - \alpha_\mu)^2,\\
E\left[\frac{\bar{s}_w^2}{n(m-1)}\right] &= \frac{1}{mn}\,\Sigma\Sigma\sigma^2_{\mu\nu} + \frac{1}{n(m-1)}\,\Sigma\Sigma(\alpha_{\mu\nu} - \bar{\alpha}_\nu)^2.
\end{aligned} \tag{7}$$

In the Bernouilli case which as far as the author knows is the only one which has been considered in this connection [5], we get the wellknown result:

$$\begin{aligned}
&E_B\left[\frac{s^2}{mn-1}\right] = E_B\left[\frac{s_a^2}{m-1}\right] = E_B\left[\frac{\bar{s}_a^2}{n-1}\right]\\
&\qquad = E_B\left[\frac{s_w^2}{m(n-1)}\right] = E_B\left[\frac{\bar{s}_w^2}{n(m-1)}\right] = E_B\left[\frac{S^2}{(m-1)(n-1)}\right].
\end{aligned} \tag{8}$$

Now let us assume a Lexis series, with

$$\alpha_{\mu\nu} = \alpha_\mu\,; \qquad \alpha_\mu \neq \alpha; \qquad \bar{\alpha}_\nu = \alpha, \qquad \sigma^2_{\mu\nu} = \sigma^2_\mu \tag{9}$$

Then (7) reduces to

$$\begin{aligned}
E_L\left[\frac{s^2}{mn-1}\right] &= \frac{1}{m}\,\Sigma\sigma^2_\mu + \frac{1}{mn-1}\,n\Sigma(\alpha_\mu - \alpha)^2,\\
E_L\left[\frac{s_a^2}{m-1}\right] &= \frac{1}{m}\,\Sigma\sigma^2_\mu + \frac{1}{m-1}\cdot n\Sigma(\alpha_\mu - \alpha)^2,\\
E_L\left[\frac{\bar{s}_a^2}{n-1}\right] &= \frac{1}{m}\,\Sigma\sigma^2_\mu\,,
\end{aligned}$$

$$E_L\left[\frac{S^2}{(m-1)(n-1)}\right] = \frac{1}{m}\,\Sigma\sigma_\mu^2,$$
$$E_L\left[\frac{s_w^2}{m(n-1)}\right] = \frac{1}{m}\,\Sigma\sigma_\mu^2,$$
$$E_L\left[\frac{\bar{s}_w^2}{n(m-1)}\right] = \frac{1}{m}\,\Sigma\sigma_\mu^2 + \frac{1}{m-1}\,\Sigma(\alpha_\mu - \alpha)^2.$$

From these formulae we deduce—besides the inequalities (11), (12) of Section 1, and the corresponding formulae where the role of rows and columns is interchanged—the further inequalities:

$$E_L\left[\frac{s_a^2}{m-1}\right] > E_L\left[\frac{\bar{s}_w^2}{n(m-1)}\right] > E_L\left[\frac{\bar{s}_a^2}{n-1}\right]. \tag{11}$$

But there are also characteristic equalities, namely:

$$E_L\left[\frac{\bar{s}_a^2}{n-1}\right] = E_L\left[\frac{S^2}{(m-1)(n-1)}\right] = E_L\left[\frac{s_w^2}{m(n-1)}\right]. \tag{12}$$

These *equalities*[9] seem often to be more appropriate than the usual *inequalities* in testing the hypothesis of a Lexis series.

Let us finally consider the Poisson case which is very often neglected. There we have:

$$\alpha_{\mu\nu} = \bar{\alpha}_\nu, \qquad \bar{\alpha}_\nu \neq \alpha, \qquad \alpha_\mu = \alpha, \qquad \sigma_{\mu\nu}^2 = \bar{\sigma}_\nu^2. \tag{13}$$

Then—beside the inequalities (13), (14) of Section 1 and the corresponding ones where the role of rows and columns is interchanged—we find the new *inequality*:

$$E_P\left[\frac{s_a^2}{m-1} - \frac{\bar{s}_a^2}{n-1}\right] = -\frac{m}{n-1}\sum_\nu (\bar{\alpha}_\nu - \alpha)^2 < 0, \tag{14}$$

which of course corresponds to the Lexis inequality (11). The characteristic *equalities* are now:

$$E_P\left[\frac{s_a^2}{m-1}\right] = E_P\left[\frac{S^2}{(m-1)(n-1)}\right] = E_P\left[\frac{\bar{s}_w^2}{n(m-1)}\right]. \tag{15}$$

These equalities (12) and (15) can be used in testing the hypothesis of Lexis or Poisson series respectively in the same way as the equalities (9) for the Bernouilli case. We shall deal with the variances of these test functions in (d) of this section.

(*b*). *Mathematical expectations of the quotients of certain test functions.* We have seen that in case of a Lexis-Series the expectations of $\frac{\bar{s}_a^2}{n-1}$, of $\frac{S^2}{(m-1)(n-1)}$ and of $\frac{s_w^2}{m(n-1)}$ are equal. We will show that even in this case

[9] See [10] pp. 81–90 for proofs of these inequalities for the case of normal populations.

$$
(16) \qquad
\begin{aligned}
E_L\left[\frac{\bar{s}_a^2}{n-1}\Big/\frac{s_w^2}{m(n-1)}\right] &= 1,\\
E_L\left[\frac{s_w^2}{m(n-1)}\Big/\frac{S^2}{(m-1)(n-1)}\right] &= 1,\\
E_L\left[\frac{\bar{s}_a^2}{n-1}\Big/\frac{S^2}{(m-1)(n-1)}\right] &= 1,\\
E_L\left[\frac{S^2}{(m-1)(n-1)}\Big/\frac{s_w^2}{m(n-1)}\right] &= 1.
\end{aligned}
$$

Let us write for the moment: $\dfrac{\bar{s}_a^2}{n-1} = \bar{T}$ and $\dfrac{S^2}{(m-1)(n-1)} = T$. As both T and $\bar{T}$ are of second degree in the $x_{\mu\nu}$ we may write:

$$\bar{T} - T = A + \sum_{\mu,\nu} B_{\mu\nu}x_{\mu\nu} + \sum_{\mu,\nu} C_{\mu\nu}x_{\mu\nu}^2 + \sum\sum D_{\mu_1,ij;\mu_2,j}x_{\mu_1,i}x_{\mu_2 j},$$

where the A, B, C, D are constants. The last sum contains $\frac{1}{2}\cdot mn(mn-1)$ terms and *not both* $\mu_1 = \mu_2$ and $i = j$ hold. Compute the expectation of $\bar{T} - T$ with respect to populations which form a Lexis series $V_{\mu\nu}(x) = V_\mu(x)$. Denote by α_μ, σ_μ^2 the respective mean values and variances. We then have because of (11):

$$0 = E_L[\bar{T} - T] = A + \sum_\mu \alpha_\mu \sum_\nu B_{\mu\nu} + \sum_\mu (\sigma_\mu^2 + \alpha_\mu^2)\sum_\nu C_{\mu\nu} + \sum_{\mu_1,\mu_2} \alpha_{\mu_1}\alpha_{\mu_2}\sum_{i,j} D_{\mu_1 i;\mu_2 j}$$

or introducing $\sum_\nu B_{\mu\nu} = B_\mu$; $\sum_\nu C_{\mu\nu} = C_\mu$; $\sum_{i,j} D_{\mu_1 i;\mu_2 j} = D_{\mu_1\mu_2}$ we get:

$$0 = E_L[\bar{T} - T] = A + \sum_\mu \alpha_\mu B_\mu + \sum_\mu (\sigma_\mu^2 + \alpha_\mu^2)C_\mu + \sum_{\mu_1,\mu_2}\alpha_{\mu_1}\alpha_{\mu_2}D_{\mu_1\mu_2}.$$

As this equality is exact for an arbitrary set of $V_\mu(x)$ we deduce that $A = 0$, $B_\mu = 0$, $C_\mu = 0$, $D_{\mu_1\mu_2} = 0$.

Let us now compute under the same assumption the expectation of $(\bar{T} - T)/T$. Here the expectations of $1/T$, $x_{\mu\nu}/T$ etc. will take the place of the expectations of 1, $x_{\mu\nu}$, $\cdots$. But these new expectations will not depend on the index ν (index within the row) because the populations are the same within each row and because of the symmetry of T in the $m\cdot n$ variables $x_{\mu\nu}$. Hence we can put

$$E\left(\frac{1}{T}\right) = l_0, \qquad E\left(\frac{x_{\mu\nu}}{T}\right) = l_\mu, \qquad E\left(\frac{x_{\mu\nu}^2}{T}\right) = l_\mu, \qquad E\left(\frac{x_{\mu_1 i}x_{\mu_2 j}}{T}\right) = l_{\mu_1\mu_2}, \quad \text{etc.}$$

and we get

$$E\left[\frac{\bar{T} - T}{T}\right] = E\left(\frac{\bar{T}}{T} - 1\right) = Al_0 + \sum_\mu l_\mu B_\mu + \sum_\mu l_\mu C_\mu + \sum_{\mu_1,\mu_2} l_{\mu_1,\mu_2}D_{\mu_1,\mu_2} = 0,$$

because all the coefficients are equal to zero. Our theorem is thus proved. The same conclusion holds if the denominator—without being symmetric in all the

$m \cdot n$ variables—does not depend on the row index. And as this last property holds for s_w^2 the expectations (16) are all shown to be equal to one.

Analogous relations are valid for Poisson series.

(*c*). *Non-independent populations.* We omit in this section the assumption of independence of the $m \cdot n$ populations but assume the theoretical population to be a general $m \cdot n$-variate distribution:

$$V(x_{11}, x_{12}, \cdots x_{mn}). \tag{17'}$$

From $V(x_{11}, x_{12}, \cdots x_{mn})$ we derive the mn one-dimensional distributions $V_{\mu\nu}(x)$ ($\mu = 1, \cdots m$; $\nu = 1, \cdots n$) by letting all the variables except $x_{\mu\nu}$ tend to $+\infty$, because $V_{\mu\nu}(x)$ is the probability that $x_{\mu\nu} \leqq x$ regardless of the values of the other variables. In a similar way we derive the $\frac{1}{2}mn(mn - 1)$ two dimensional distributions $V_{\mu_1\nu_1;\mu_2\nu_2}(x, y)$, that is the probability that $x_{\mu_1\nu_1} \leqq x$ and $x_{\mu_2\nu_2} \leqq y$. We get this distribution from (17′) as all the variables with the exception of $x_{\mu_1\nu_1}$ and $x_{\mu_2\nu_2}$ tend to $+\infty$. We denote as before by $\alpha_{\mu\nu}$ and $\sigma_{\mu\nu}^2$ the expectation of $x_{\mu\nu}$ and $(x_{\mu\nu} - \alpha_{\mu\nu})^2$ respectively. But the expectation of $(x_{\mu_1\nu_1} - \alpha_{\mu_1\nu_1})(x_{\mu_2\nu_2} - \alpha_{\mu_2\nu_2})$ which was zero in case of the independence of $x_{\mu_1\nu_1}$ and $x_{\mu_2\nu_2}$ may now differ from zero. Denote by $\mathfrak{E}$ the expectation with respect to (17′). Then:

$$\begin{aligned} &\mathfrak{E}[(x_{\mu_1\nu_1} - \alpha_{\mu_1\nu_1})(x_{\mu_2\nu_2} - \alpha_{\mu_2\nu_2})] \\ &\quad = \int\int \cdots \int (x_{\mu_1\nu_1} - \alpha_{\mu_1\nu_1})(x_{\mu_2\nu_2} - \alpha_{\mu_2\nu_2})\, dV(x_{11}, \cdots x_{mn}) \\ &\quad = \int\int (x - \alpha_{\mu_1\nu_1})(y - \alpha_{\mu_2\nu_2})\, dV_{\mu_1\nu_1;\nu_2\mu_2}(xy) = R_{\mu_1\nu_1;\mu_2\nu_2} = R_{\mu_2\nu_2;\mu_1\nu_1}. \end{aligned} \tag{17}$$

Let us first deduce a general formula for the expectation of a sample variance in the case of dependent populations. Let $P(y_1, \cdots y_r)$ be the distribution of r chance variables $y_1, \cdots y_r$ which have the average b. Denoting by β_ρ the expectation of y_ρ with respect to P, by β the average of the β_ρ, by τ_ρ^2 the expectation of $(y_\rho - \beta_\rho)^2$ by R_{ij} that of $(y_i - \beta_i)(y_j - \beta_j)$ we find, without difficulty, for the expectation of the sample variance

$$\begin{aligned} &\text{Exp.}\left[\frac{1}{r}\sum_{\rho=1}^{r}(y_\rho - b)^2\right] \\ &\quad = \frac{1}{r}\int \cdots \int [(y_1 - b)^2 + \cdots + (y_r - b)^2]\, dP(y_1, \cdots y_r) \\ &\quad = \frac{r-1}{r^2}\sum_{\rho=1}^{r}\tau_\rho^2 + \frac{1}{r}\sum_{\rho}(\beta_\rho - \beta)^2 - \frac{2}{r^2}\sum_{i<j} R_{ij}. \end{aligned} \tag{18}$$

Let us apply this result in the computation of the expectations of our test functions. It is not difficult to compute them in the general case of *different* mean values and variances. But we restrict ourselves to the consideration of certain particular cases. Take first the case where all the $m \cdot n$ mean values $\alpha_{\mu\nu}$ are equal

to each other and likewise the $m \cdot n$ variances and the $\frac{1}{2}mn(mn - 1)$ covariances. Denote these magnitudes by α, σ^2 and R, respectively, we see from (18) that:

$$\mathcal{E}\left(\frac{s^2}{mn-1}\right) = \mathcal{E}\left(\frac{s_a^2}{m-1}\right) = \mathcal{E}\left(\frac{\bar{s}_a^2}{n-1}\right)$$

$$(19) \qquad = \mathcal{E}\left(\frac{s_w^2}{m(n-1)}\right) = \mathcal{E}\left(\frac{\bar{s}_w^2}{n(m-1)}\right) = \mathcal{E}\left(\frac{S^2}{(m-1)(n-1)}\right)$$

$$= \sigma^2 - R.$$

We have thus obtained the result *that in the case of dependent populations, just described, the expectations of the six different test functions are still the same.*

Of course we may assume many other particular kinds of mutual dependence of the populations. The following assumption seems to be appropriate for problems where rows and columns play a *different* role: We consider dependence *only within each* row, that means we assume only the variables $x_{\mu 1}, x_{\mu 2}, \cdots x_{\mu n}$ as mutually dependent. The distribution (16) has then the following form:

$$(20) \qquad V(x_{11}, \cdots x_{mn}) = V_1(x_{11}, \cdots x_{1n})V_2(x_{21}, \cdots x_{2n}) \cdots V_m(x_{m1}, \cdots x_{mn}).$$

In the usual way we derive the $m \cdot n$ one dimensional distributions $V_{\mu\nu}(x)$ and the $\frac{1}{2}mn(mn - 1)$ two-dimensional distributions $V_{\mu_1\nu_1;\mu_2\nu_2}(x, y)$. If $\mu_1 \neq \mu_2$ such a two-dimensional distribution reduces to the product of the respective one-dimensional distributions. Only the $\frac{1}{2}mn(n - 1)$ bivariate distributions derived from one and the same $V_\mu(x_{\mu 1}, \cdots x_{\mu n})$ will not reduce in this way.

Denoting again by $\mathcal{E}$ the expectation with respect to $V(x_{11}, \cdots x_{mn})$ we find:

$$(21) \qquad \begin{aligned} \mathcal{E}[(x_{\mu_1 i} - \alpha_{\mu_1 i})(x_{\mu_2 j} - \alpha_{\mu_2 j})] &= 0 && \mu_1 \neq \mu_2 \\ &= R_{ij}^{(\mu_1)} && \mu_1 = \mu_2 \text{ and } i \neq j. \end{aligned}$$

Applying now formula (18) in the computation of the expectations of s^2, s_w^2 and s_a^2 we find:

$$\mathcal{E}[\sum\sum (x_{\mu\nu} - a)^2] = \frac{mn-1}{mn}\sum\sum \sigma_{\mu\nu}^2 + \sum\sum (\alpha_{\mu\nu} - \alpha)^2 - \frac{2}{mn}\sum_{\mu=1}^{m}\sum_{i<j} R_{ij}^{(\mu)},$$

$$(22) \qquad \mathcal{E}[\sum\sum (x_{\mu\nu} - a_\mu)^2] = \frac{m(n-1)}{mn}\sum\sum \sigma_{\mu\nu}^2 + \sum\sum (\alpha_{\mu\nu} - \alpha_\mu)^2 - \frac{2}{n}\sum_{\mu=1}^{m}\sum_{i<j} R_{ij}^{(\mu)},$$

$$\mathcal{E}[\sum\sum (a_\mu - a)^2] = \frac{m-1}{mn}\sum\sum \sigma_{\mu\nu}^2 + n\sum (\alpha_\mu - \alpha)^2 + \frac{2(m-1)}{mn}\sum_{\mu=1}^{m}\sum_{i<j} R_{ij}^{(\mu)}.$$

Let us now suppose that *all the $m \cdot n$ distributions are equal* to each other, or, at least, that:

$$\alpha_{\mu\nu} = \alpha. \tag{23}$$

This assumption, which is characterized by (21), is, of course, different from the one which leads us to (19). We find now by means of (22), if we set

$$\sum_{\mu=1}^{m} \sum_{i<j} R_{ij}^{\mu} = \bar{R} \quad \text{and} \quad \frac{2}{mn(mn-1)}\bar{R} = R,$$

$$\mathcal{E}\left[\frac{s_a^2}{m-1} - \frac{s^2}{mn-1}\right] = \frac{2}{mn-1}\bar{R} = \frac{mn(n-1)}{mn-1}R. \tag{24}$$

Assuming $R > 0$ (positive average correlation) we may compare this result with (11) Section 1. The term on the right side of (24) is also of the same order of magnitude as that in (11).—For negative R the term on the right side of (24) is negative and the equation may be compared with (13) Section 1. We see that for the test functions s^2/r and s_a^2/r_a "*positive, (negative) average correlation within rows" has the same effect as "Lexis (Poisson) Series" of populations.*

Consider now the test functions $\bar{s}_a^2$ and S^2. We find

$$\mathcal{E}[\bar{s}a^2] = \mathcal{E}[\Sigma\Sigma(\bar{a}_\nu - a)^2] = \frac{n-1}{mn}\Sigma\Sigma\sigma_{\mu\nu}^2 + m\Sigma(\bar{\alpha}_\nu - \alpha)^2 - \frac{2}{mn}\bar{R}, \tag{25}$$

and

$$\begin{aligned}\mathcal{E}[S^2] = \mathcal{E}[\Sigma\Sigma(x_{\mu\nu} - a_\mu - \bar{a}_\nu + a)^2] &= \frac{(m-1)(n-1)}{mn}\Sigma\Sigma\sigma_{\mu\nu}^2 \\ &+ \Sigma\Sigma(\alpha_{\mu\nu} - \alpha_\mu - \bar{\alpha}_\nu + \alpha)^2 - \frac{2(m-1)}{mn}\bar{R}.\end{aligned} \tag{25'}$$

Assuming (23) we get:

$$\mathcal{E}\left[\frac{s_a^2}{m-1} - \frac{\bar{s}_a^2}{n-1}\right] = nR, \tag{26}$$

and if $R > 0$:

$$\mathcal{E}\left[\frac{s_a^2}{m-1}\right] > \mathcal{E}\left[\frac{\bar{s}_w^2}{n(m-1)}\right] > \mathcal{E}\left[\frac{\bar{s}_a^2}{n-1}\right]. \tag{26'}$$

The first equality is analogous to (11) and (14) of Section 2 for positive or negative R respectively.[10] We also get under the assumption (23)

$$\mathcal{E}\left[\frac{\bar{s}_a^2}{n-1}\right] = \mathcal{E}\left[\frac{S^2}{(m-1)(n-1)}\right] = \mathcal{E}\left[\frac{s_w^2}{m(n-1)}\right]. \tag{27}$$

[10] I have studied in another paper the *combination* of Lexis series and "positive correlation within rows." It turns out that the two kinds of positive effects reinforce each other. The same is true for "negative correlation" and Poisson series. See [3].

These are the same equations as (12) Section 2, and they are true for either sign of R. Hence they provide no way to decide between Lexis series and correlated populations. But computing the expectations of the magnitudes which occur in (15) Section 2 we find from (22), (25) and (25′)

$$\mathcal{E}\left[\frac{s_a^2}{m-1}\right] = \sigma^2 + (n-1)R, \qquad \mathcal{E}\left[\frac{\bar{s}_w^2}{n(m-1)}\right] = \sigma^2 \tag{28}$$
$$\mathcal{E}\left[\frac{S^2}{(m-1)(n-1)}\right] = \sigma^2 - R.$$

And hence we may say:
If the observed value of $s_a^2/(m-1)$ is greater than that of $\bar{s}_w^2/n(m-1)$ this can be explained either by the assumption of a Lexis series or a positive correlation within rows; but their equality indicate, a Poisson series; and if the first is smaller than the second we may assume negative correlation.
In the same way we may explain

$$\left[\frac{\bar{s}_w^2}{n(m-1)}\right]_{\text{observed}} > \left[\frac{S^2}{(m-1)(n-1)}\right]_{\text{observed}},$$

either by positive correlation or by Lexis series; whereas the equality indicates a Poisson series and the sign $<$ indicates negative correlation.

(*d*). *The variances of the test functions.* We have still to find the variance of our test functions. Let us compute the variance of

$$\Sigma\Sigma(x_{\mu\nu} - a_\mu - \bar{a}_\nu + a)^2$$

with respect to the $m \cdot n$ dimensional distribution $V(x_{11})V(x_{12}) \cdots V(x_{mn})$. Let us put

$$x_{\mu\nu} - a_\mu - \bar{a}_\nu + a = y_{\mu\nu}, \tag{29}$$

then we see that the average of the $y_{\mu\nu}$ equals zero

$$\bar{y} = \frac{1}{mn}\Sigma\Sigma y_{\mu\nu} = a - \frac{1}{mn} n\Sigma a_\mu - \frac{1}{mn} m\Sigma\bar{a}_\nu + a = 0.$$

and

$$S^2 = \Sigma\Sigma(x_{\mu\nu} - a_\mu - \bar{a}_\nu + a)^2 = \Sigma\Sigma(y_{\mu\nu} - \bar{y})^2.$$

Each $y_{\mu\nu}$ is a linear function of the $x_{\mu\nu}$ e.g.

$$\begin{aligned} y_{11} &= x_{11}\frac{(m-1)(n-1)}{mn} \\ &\quad - \frac{m-1}{mn}\sum_{j=2}^{n} x_{1j} - \frac{n-1}{mn}\sum_{i=2}^{m} x_{i1} + \frac{1}{mn}\sum_{2}^{m}\sum_{2}^{n} x_{ij} \\ &= x_{11}\lambda_2 + \lambda_2\sum_{2}^{n} x_{1j} + \lambda_3\sum_{2}^{m} x_{i1} + \lambda_4\sum_{2}^{m}\sum_{2}^{n} x_{ij}. \end{aligned} \tag{30}$$

Using the same notations as in Section 1 (c) we find, because of the independence of each chance variable

$$\begin{aligned}\text{Var}\,(y_{11}) &= \lambda_1^2\sigma^2 + \lambda_2^2(n-1)\sigma^2 + \lambda_3^2(m-1)\sigma^2 \\ &\quad + \lambda_4^2(m-1)(n-1)\sigma^2 = \frac{(m-1)(n-1)}{mn}\sigma^2\end{aligned} \tag{31'}$$

and we find the same result for each $y_{\mu\nu}$:

$$\sigma'^2 = \text{Var}\,(y_{\mu\nu}) = \frac{(m-1)(n-1)}{mn}\sigma^2 \tag{31}$$

in agreement with the fourth line of (7) of this section. We still need M_4' the fourth moment about the mean of $y_{\mu\nu}$ which we can compute from the fourth moment of a sum. We find

$$M_4' = AM_4 + 6B\sigma^4, \tag{32}$$

and we have

$$\begin{aligned}A &= \lambda_1^4 + (n-1)\lambda_2^4 + (m-1)\lambda_3^4 + (m-1)(n-1)\lambda_4^4 \\ &= \frac{(m-1)(n-1)}{m^3n^3}(m^2-3m+3)(n^2-3n+3),\end{aligned} \tag{33}$$

and

$$\begin{aligned}B &= \lambda_1^2\{\lambda_2^2(n-1) + \lambda_3^2(m-1) + \lambda_4^2(m-1)(n-1)\} \\ &\quad + \lambda_2^2(n-1)\{\tfrac{1}{2}\lambda_2^2(n-2) + \lambda_3^2(m-1) + \lambda_4^2(m-1)(n-1)\} \\ &\quad + \lambda_3^2(m-1)\{\tfrac{1}{2}\lambda_3^2(m-2) + \lambda_4^2(m-1)(n-1)\} \\ &\quad + \tfrac{1}{2}\lambda_4^4(m-1)(n-1)[(m-1)(n-1)-1].\end{aligned} \tag{34'}$$

If we introduce the values of λ_1, λ_2, λ_3, λ_4 we find

$$\begin{aligned}m^4n^4B &= (m-1)^3(n-1)^3(m+n) + (m-1)^2(n-1)^2(m+n-2) \\ &\quad + \tfrac{1}{2}(m-1)(n-1)[(m-1)^3(n-2) + (n-1)^3(m-2) \\ &\quad + (mn-m-n)]\end{aligned} \tag{34}$$

this expression as well as that of A may be easily computed for different values of m and n.

If m and n are large, B is of order $\frac{1}{m} + \frac{1}{n}$; from (31)–(34) we see that in this case σ'^2 is approximately equal to σ^2 and M_4' to M_4.

Using now (18′) we find finally

$$\text{Var}\,\{\Sigma\Sigma\,(x_{\mu\nu} - a_\mu - \bar{a}_\nu + a)^2\} = \frac{mn-1}{mn}\{(mn-1)M_4' - (mn-3)\sigma'^4\}$$

where M_4' and σ'^2 are the expressions just computed. If we compare the variances of the test functions $s_a^2/(m-1)$ and $S^2/(m-1)(n-1)$ we see that whereas the variance of the first expression is of order $1/m$ that of the second is of order $1/mn$. Hence for large values of n the *latter expression is more exact* than the former (see the analogous remark Section 1 (c)). A similar statement can be made if $\bar{s}_a^2/(n-1)$ takes the place of $s_a^2/(m-1)$.

3. Bivariate distributions. Analysis of covariance.

(*a*). *Problem.* Suppose m persons are throwing two dice, n times; we observe the respective numbers on each die in these $m \cdot n$ trials. Or we observe on m groups of n persons the color of the hair and of the eyes. Or else we state for n years the yield of wheat (in bushels) per acre and the production cost (per bushel) for m farms; etc.

We consider $m \cdot n$ pairs of numbers $x_{\mu\nu}$, $y_{\mu\nu}$. Let $V_{\mu\nu}(x, y)$[11] be the probability that $x_{\mu\nu} \leqq x$ and $y_{\mu\nu} \leqq y$; $V_{\mu\nu}(x, +\infty) = V_{\mu\nu}^{(1)}(x)$, $V_{\mu\nu}(+\infty, y) = V_{\mu\nu}^{(2)}(y)$ and introduce the following mean values and variances

$$\text{(1)} \qquad \iint x\, dV_{\mu\nu}(x, y) = \alpha_{\mu\nu}, \qquad \iint y\, dV_{\mu\nu}(x, y) = \beta_{\mu\nu},$$

$$\text{(2)} \qquad \iint (x - \alpha_{\mu\nu})^2\, dV_{\mu\nu}(x, y) = \sigma_{\mu\nu}^2, \qquad \iint (y - \beta_{\mu\nu})^2\, dV_{\mu\nu}(xy) = \tau_{\mu\nu}^2,$$

$$\text{(3)} \qquad \iint (x - \alpha_{\mu\nu})(y - \beta_{\mu\nu})\, dV_{\mu\nu}(x, y) = \gamma_{\mu\nu}$$

$$\text{(4)} \qquad \begin{aligned} &\frac{1}{n}\sum_{\nu} \alpha_{\mu\nu} = \alpha_\mu, \qquad \frac{1}{m}\sum_{\mu} \alpha_{\mu\nu} = \bar{\alpha}_\nu, \qquad \frac{1}{mn}\sum\sum \alpha_{\mu\nu} = \alpha \\ &\frac{1}{n}\sum_{\nu} \beta_{\mu\nu} = \beta_\mu, \qquad \frac{1}{m}\sum_{\mu} \beta_{\mu\nu} = \bar{\beta}_\nu, \qquad \frac{1}{mn}\sum\sum \beta_{\mu\nu} = \beta \end{aligned}$$

Let us compute the mathematical expectations of certain test functions with respect to the $2mn$-dimensional distributions

$$V_{11}(x_{11}, y_{11})V_{12}(x_{12}, y_{12}) \cdots V_{mn}(x_{mn}, y_{mn}). \quad \text{Let}$$

$$\text{(5)} \qquad \begin{aligned} &E[F(x_{11}, y_{11}, \cdots x_{mn}, y_{mn})] \\ &\qquad = \int \cdots \int F(x_{11}, \cdots y_{mn})\, dV_{11}(x_{11}, y_{11}) \cdots dV_{mn}(x_{mn} y_{mn}) \end{aligned}$$

[11] In the particular case where $V_{\mu\nu}(x, y)$ has everywhere a derivative $\dfrac{\partial^2 V_{\mu\nu}}{\partial x\, \partial y}$ we can use the two dimensional density $v_{\mu\nu}(x, y) = \dfrac{\partial^2 V_{\mu\nu}}{\partial x\, \partial y}$ and the one-dimensional densities

$$v_{\mu\nu}^{(1)}(x) = \int v_{\mu\nu}(xy)\, dy; \qquad v_{\mu\nu}^{(2)}(y) = \int v_{\mu\nu}(x, y)\, dx$$

and we have

$$V_{\mu\nu}^{(1)}(x) = \int_{-\infty}^{x} v_{\mu\nu}^{(1)}(x)\, dx, \qquad V_{\mu\nu}^{(2)}(y) = \int_{-\infty}^{y} v_{\mu\nu}^{(2)}(y)\, dy.$$

We then have[12]

$$(5') \quad F[G(x_{11}, \cdots x_{mn})] = \int_{(x_{11})} \cdots \int_{(x_{mn})} G(x_{11} \cdots x_{mn})\, dV_{11}^{(1)}(x_{11}) \cdots dV_{mn}^{(1)}(x_{mn}).$$

In analogy with previous notations we introduce

$$(6) \quad \begin{aligned} a_\mu &= \frac{1}{n} \sum_\nu x_{\mu\nu}, & \bar{a}_\nu &= \frac{1}{m} \sum_\mu x_{\mu\nu}, & a &= \frac{1}{mn} \sum \sum x_{\mu\nu}, \\ b_\mu &= \frac{1}{n} \sum_\nu y_{\mu\nu}, & \bar{b}_\nu &= \frac{1}{m} \sum_\mu y_{\mu\nu}, & b &= \frac{1}{mn} \sum \sum y_{\mu\nu}, \end{aligned}$$

and

$$(7) \quad \begin{aligned} s^2 &= \Sigma\Sigma(x_{\mu\nu} - a)^2, & s_a^2 &= n\Sigma(a_\mu - a)^2, & s_w^2 &= \Sigma\Sigma(x_{\mu\nu} - a_\mu)^2 \\ S^2 &= \Sigma\Sigma(x_{\mu\nu} - a_\mu - \bar{a}_\nu + a)^2, & \bar{s}_a^2 &= m\Sigma(\bar{a}_\nu - a)^2, & \bar{s}_w^2 &= \Sigma\Sigma(x_{\mu\nu} - \bar{a}_\nu)^2 \\ t^2 &= \Sigma\Sigma(y_{\mu\nu} - b)^2, & t_a^2 &= n\Sigma(b_\mu - b)^2, & t_w^2 &= \Sigma\Sigma(y_{\mu\nu} - b_\mu)^2 \\ T^2 &= \Sigma\Sigma(y_{\mu\nu} - b_\mu - \bar{a}_\nu + b)^2, & \bar{t}_a^2 &= m\Sigma(\bar{b}_\nu - b)^2, & \bar{t}_w^2 &= \Sigma\Sigma(y_{\mu\nu} - \bar{b}_\nu)^2, \end{aligned}$$

and

$$(8) \quad \begin{aligned} c &= \Sigma\Sigma(x_{\mu\nu} - a)(y_{\mu\nu} - b), & C &= \Sigma\Sigma(x_{\mu\nu} - a_\mu - \bar{a}_\nu + a)(y_{\mu\nu} - b_\mu - \bar{b}_\nu + b) \\ c_a &= n\Sigma(a_\mu - a)(b_\mu - b) & c_w &= \Sigma\Sigma(x_{\mu\nu} - a_\mu)(y_{\mu\nu} - b_\mu) \\ \bar{c}_a &= m\Sigma(\bar{a}_\nu - a)(\bar{b}_\nu - b) & \bar{c}_w &= \Sigma\Sigma(x_{\mu\nu} - \bar{a}_\nu)(y_{\mu\nu} - \bar{b}_\nu) \end{aligned}$$

we then have

$$(9) \quad \begin{aligned} s^2 &= S^2 + s_a^2 + \bar{s}_a^2 = s_w^2 + s_a^2 = \bar{s}_w^2 + \bar{s}_a^2, \\ t^2 &= T^2 + t_a^2 + \bar{t}_a^2 = t_w^2 + t_a^2 = \bar{t}_w^2 + \bar{t}_a^2, \\ c &= C + c_a + \bar{c}_a = c_a + c_w = \bar{c}_a + \bar{c}_w, \end{aligned}$$

and corresponding relations for the ranks of these quadratic forms. We find for the expectations of these test functions, in analogy with previously investigated formulae:

$$E\left[\frac{t^2}{mn - 1}\right] = \frac{1}{mn} \Sigma\Sigma\tau_{\mu\nu}^2 + \frac{1}{mn - 1} \Sigma\Sigma(\beta_{\mu\nu} - \beta)^2,$$

$$E\left[\frac{t_a^2}{m - 1}\right] = \frac{1}{mn} \Sigma\Sigma\tau_{\mu\nu}^2 + \frac{1}{m - 1} n\Sigma(\beta_\mu - \beta)^2,$$

. . . .

and

$$E\left[\frac{c}{mn - 1}\right] = \frac{1}{mn} \Sigma\Sigma\gamma_{\mu\nu} + \frac{1}{mn - 1} \Sigma\Sigma(\alpha_{\mu\nu} - \alpha)(\beta_{\mu\nu} - \beta),$$

$$E\left[\frac{c_a}{m - 1}\right] = \frac{1}{mn} \Sigma\Sigma\gamma_{\mu\nu} + \frac{1}{m - 1} \cdot n\Sigma(\alpha_\mu - \alpha)(\beta_{\mu\nu} - \beta),$$

. . . .

[12] It may be mentioned that the problem considered in this section of mn bivariate distribution $v_{\mu\nu}(x, y)$ constitutes, of course, only a particular case of dependence (see section 2, (c)) for a $2mn$ dimensional population $v(x_{11}, y_{11}, x_{12}, y_{12}, \cdots x_{mn}, y_{mn})$.

1) *If all the $\alpha_{\mu\nu}$ equal each other, or all the $\beta_{\mu\nu}$ equal each other, we find*:

$$E_B\left[\frac{c}{mn-1}\right] = E_B\left[\frac{c_a}{m-1}\right] = E_B\left[\frac{c_w}{m(n-1)}\right]$$

$$= E_B\left[\frac{C}{(m-1)(n-1)}\right] = E_B\left[\frac{\bar{c}_a}{n-1}\right] = E_B\left[\frac{\bar{c}_w}{n(m-1)}\right] = \frac{1}{mn}\Sigma\Sigma\gamma_{\mu\nu}.$$

These formulae provide us with unbiased estimates of $\Sigma\Sigma\gamma_{\mu\nu}$.

2) *The $\alpha_{\mu\nu}$ are equal within each row but differ from row to row,* (Lexis) $\alpha_{\mu\nu} = \alpha_\mu \neq \alpha$; $\bar{\alpha}_\nu = \alpha$ *whereas the $\beta_{\mu\nu}$ may have arbitrary values, then*

$$E_L\left[\frac{\bar{c}_a}{n-1}\right] = E_L\left[\frac{c_w}{m(n-1)}\right] = E_L\left[\frac{C}{(m-1)(n-1)}\right]. \tag{13}$$

The same equalities are valid for arbitrary $\alpha_{\mu\nu}$ if the $\beta_{\mu\nu} = \beta_\mu$; $\bar{\beta}_\nu = \beta$. Our new equalities may be of some interest because inequalities analogous to those of the Lexis case cannot be proved for covariances. If the observed values of the expressions in (13) are significantly different we may conclude that neither the $\alpha_{\mu\nu}$ nor the $\beta_{\mu\nu}$ form a Lexis series. A judgment of the test (13) might be based on the investigation of its power function. But besides we have the equalities (12) and analogous equalities containing $\bar{t}_a^2$, T^2 and t_w^2.

3) *If either* $\alpha_{\mu\nu} = \bar{\alpha}_\nu$, $\bar{\alpha}_\nu \neq \alpha$, $\alpha_\mu = \alpha$,

or $\beta_{\mu\nu} = \bar{\beta}_\nu$, $\beta_\nu \neq \beta$, $\beta_\mu = \beta$.

We have the new equalities

$$E_P\left[\frac{c_a}{m-1}\right] = E_P\left[\frac{\bar{c}_w}{n(m-1)}\right] = E_P\left[\frac{C}{(m-1)(n-1)}\right], \tag{14}$$

and there are no inequalities analogous to the inequalities (14) *of Section* 2, *and* (13), (14) *of Section* 1.

Most of the investigations of Sections 1 and 2 can be generalized for this two dimensional problem.

BIBLIOGRAPHY

[1] R. A. FISHER, *Statistical Methods for Research Workers*, 6th ed., p. 214 ff.

[2] R. A. FISHER, "Applications of 'Student's' distributions," *Metron*, Vol. 5 (1926), pp. 90–104.

[3] H. GEIRINGER, "A new explanation of non-normal dispersion in the Lexis theory," *Econometrica*, Vol. 10 (1942), pp. 53–60.

[4] F. R. HELMERT, *Zeits. für Math. und Physik*, Vol. 21 (1876), p. 192–218.

[5] I. O. IRWIN, "Mathematical theorems involved in the analysis of variance," *Jour. Roy. Stat. Soc.*, Vol. 94 (1931), pp. 284–300.

[6] W. G. MADOW, "Limiting distributions of quadratic and bilinear forms," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 125–147.

[7] R. v. MISES, "Theorie des probabilites. Fondements et applications," *Annales de l'Institut Poincare*, (1931), pp. 137–190.

[8] H. L. Rietz, "On the Lexis theory and the analysis of variance," *Bull. Am. Math. Soc.*, (1932), pp. 731 ff.
[9] A. A. Tschuprow, *Skandinavisk Aktuarietidskrift*, Vol. 6 (1918).
[10] A. Wald, *Lectures on the Analysis of Variance and Covariance*, *Columbia University*, 1941.
[11] Milton Friedman, "The use of ranks to avoid the assumption of normality," *Jour. Amer. Stat. Assn.*, Vol. 32 (1937), pp. 675–701.